Technology
Assessment
Program

Office of Patient Care Services

UPDATED INFORMATION FOR VA TECHNOLOGY ASSESSMENT PROGRAM (VATAP) REPORTS

In June 2000, VATAP was relocated within the Veterans Health Administration from the Office of Research & Development to the Office of Patient Care Services. The following report was produced prior to the relocation of VATAP.

Current VATAP contact information is as follows:

VA Technology Assessment Program (11T)

VA Boston Healthcare System

150 South Huntington Avenue

Boston, MA 02130

Tel: 857.364.4469 Fax: 857.364.6587
vatap@med.va.gov
http://www.va.gov/vatap http://vaww.va.gov/vatap

Technology Assessment Program Report

Positron Emission Tomography:

- Descriptive Analysis of Experience with PET in VA
- Systematic Reviews: FDG-PET as a Diagnostic Test for Cancer and Alzheimer's Disease

October, 1996

Primary Authors:
Karen Flynn, D.D.S., M.S., *Manager*
Elizabeth Adams, R.R.T., M.P.H., *Management & Program Analyst*

Contributing Author:
Diana Anderson, R.N., M.P.H., *Research Analyst*
Technology Assessment Program

Management Decision and Research Center
Health Services Research and Development Service
Office of Research and Development
Department of Veterans Affairs

Karen Flynn, D.D.S., M.S., Manager
Technology Assessment Program
Management Decision and Research Center
VA Medical Center (152M)
150 South Huntington Avenue
Boston, MA 02130

Tel: (617) 278-4469   FTS: (700) 839-4469   Fax: (617) 278-4438
flynn.karen@forum.va.gov
g.mdrc-ta@forum.va.gov

TABLE OF CONTENTS

Executive Summary

Appendix 1 Advisory Committee to the PET Assessment

Appendix 2 Assessing Diagnostic Technologies

Appendix 3 Systematic review: PET as a Diagnostic Test in Head and Neck Cancer

Appendix 4 Systematic review: PET as a Diagnostic Test in Colorectal Cancer

Appendix 5 Systematic review: PET as a Diagnostic Test in Breast Cancer

Appendix 6 Systematic review: PET as a Diagnostic Test in Lung Cancer

Appendix 7 Systematic review: PET as a Diagnostic Test in Solitary Pulmonary Nodules

Appendix 8 Systematic review: PET as a Diagnostic Test in Alzheimer's Disease

Appendix 9 Experience With PET in VHA

Acknowledgements

The MDRC Technology Assessment Program wishes to thank the members of the Advisory Committee to the PET Assessment, who contributed to the design of the assessment, confirmed the completeness of literature retrieval, reviewed the report, and reached consensus on its recommendations.

Chair: *Marguerite Hays, M.D.*, ACOS Research Service, Palo Alto VA Medical Center

Members: *Martin Charns, D.B.A.*, Director, Management Decision and Research Center

Daniel Deykin, M.D., Director, Health Services Research and Development Service

James Fletcher, M.D., Chief, Nuclear Medicine Service, St. Louis VA Medical Center

Clifford Goodman, Ph.D., Consultant

Milton Gross, M.D., Program Director, VA Nuclear Medicine Service

Thomas Holohan, M.D., Director, Office of Health Technology Assessment, Agency for Health Care Policy and Research

Steven Hotta, M.D., Ph.D., Project Manager for PET, Office of Health Technology Assessment, Agency for Health Care Policy and Research

Steven Larson, M.D., Chief, Nuclear Medicine Service, Memorial Sloane Kettering Hospital

H. William Strauss, M.D., Division of Nuclear Medicine, Stanford University School of Medicine

The MDRC Technology Assessment Program thanks the physicians who suggested improvements to the disease-specific systematic reviews and other sections of the report, and acknowledges their contributions:

Alan Garber, M.D., Ph.D., Associate Director of Health Economics and Technology Assessment, Center for Health Care Evaluation, VA Palo Alto Health Care System	Methods
Thomas Holohan, M.D., Chief Patient Care Services Officer, Department of Veterans Affairs	Head and neck, colorectal, and breast cancer
L. Jack Faling, M.D., Associate Chief, Medical Service, VA Medical Center, Boston *Charles Powell, M.D.*, Staff Physician, Boston University Pulmonary Center and VA Medical Center, Boston	Lung cancer Solitary pulmonary nodules
John Booss, M.D., Director of Neurology, Department of Veterans Affairs *Thomas Bird, M.D.*, Chief, Neurology Service, Seattle VA Medical Center, and Professor of Neurology, University of Washington *Jeffrey Cummings, M.D.*, Professor of Neurology and Psychiatry, University of California at Los Angeles, and Neurology Service, West Los Angeles VA Medical Center *Judith Salerno, M.D.*, Chief Consultant for Geriatrics and Extended Care, Department of Veterans Affairs	Alzheimer's disease

The MDRC Technology Assessment Program also wishes to thank Lois Camberg, Ph.D., Mark Prashker, M.D., M.P.H., Elizabeth Kramer, and the many VA and university affiliate physicians and other staff who participated in the interviews, site visits to PET centers, and surveys.

Technology Assessment: Positron Emission Tomography

Executive Summary

The Veterans Health Administration (VHA) shares, with some of its academic affiliates, the ownership and operation of 10 positron emission tomography (PET) imaging facilities. Significant resource commitments are associated with the acquisition, maintenance, and ongoing operation of these facilities. In late 1993, the Acting Under Secretary for Health, Department of Veterans Affairs (VA), requested that the Management Decision and Research Center (within the Health Services Research and Development Service) conduct an assessment of PET. The assessment would supply the Under Secretary with information that would assist in setting future VHA policy regarding PET. The Acting Under Secretary asked two questions:

> ***What is known about the utilization of PET, and other experience with the technology, in VHA today?***
>
> ***Should VHA establish additional PET centers?***

This document reports the results of the assessment. The overall approach and findings of the assessment are presented in this summary section. The appendices detail the individual components of the assessment and provide background to the development of the assessment methodology.

The Technology Assessment Program of the Management Decision and Research Center (MDRC) focuses on evaluating the clinical applications (rather than the technical performance or technical specifications) of health care technologies, using systematic reviews of published evidence supplemented by primary data collection. The Program uses the broad definition of health care technology developed in 1978 by the Office of Technology Assessment:

> *"... the drugs, devices, and medical and surgical procedures used in health care, and the organizational and supportive systems within which such care is delivered."*

and the Institute of Medicine's 1985 definition of technology assessment:

> *"...any process of examining and reporting properties of a medical technology used in health care, such as safety, efficacy, feasibility, and indications for use, cost, and cost-effectiveness, as well as social, economic, and ethical consequences, whether intended or unintended."*

The purpose of technology assessment is to inform technology-related policy making in health care.

I. BACKGROUND

Positron emission tomography (PET) is a nuclear medicine technology that allows the visualization and measurement of biochemical processes within tissues. PET's particular functional imaging capacity is related to the physics of the positron emission detection method and to the variety of radiolabelled compounds that can be used.

Nuclear medicine imaging techniques rely on the detection of photons produced from the decay of radioactive isotopes attached to tracers that target physiologic processes (Gritters and Wahl, 1993). PET, like other nuclear medicine techniques, makes it possible to measure local tissue and organ function, re-defining disease in terms of quantifiably abnormal regional chemistry. PET and other nuclear medicine imaging therefore may complement the information obtained from other imaging methods, such as radiography, computed tomography (CT), or magnetic resonance imaging (MRI), which rely on predominantly anatomic definitions of disease (Maisey and Jeffery, 1991).

Traditional nuclear imaging is based on photon detection using a stationary single or double-headed gamma camera that produces two-dimensional images. Tomographic techniques (single photon emission tomography, SPECT) may use mechanically rotating camera heads to acquire many pictures in a 360^{o} circle around the patient. The imaging data are then reconstructed to produce multiple cross-sectional images.

Most radioactive isotopes with potential uses in medical imaging decay by releasing energy as single gamma rays (photons) whose energies fall within a range from 80 to 400 KeV. The relatively low energy of the photon released during SPECT imaging means that attenuation and scatter by tissues can degrade the image.

The radioactive isotopes used in PET decay by other means: they emit a positively charged electron (positron) from the nucleus. The positron usually travels only a very short distance (1 to 2 millimeters) before colliding with a local electron. The collision results in the annihilation of the mass of the two particles, and the emission of two gamma rays (photons) of high energy (511 KeV), which travel out at approximately 180^{o} from each other. Radiation detectors in a PET camera, which are arranged in a ring around the patient, detect the two gamma rays from each such collision simultaneously. The exact site of origin of each signal is recorded, and a cross-sectional image is displayed.

The high energy of the photon released during PET imaging means that very little of that energy is attenuated or scattered by tissue. Other sources of scatter are minimized by coincidence counting (the recording of only those photons which have been emitted at 180^{o} from each other and hit opposing crystals in the camera simultaneously).

All medical imaging involves comparisons: of an image with the interpreter's mental pictures of the patterns representative of "normal" and of different disease states; or of changes in sequential images from the same patient (Links and Devous, 1994). PET and other nuclear medicine image patterns represent spatial and temporal arrangements and rearrangements of the physiological or biochemical process under investigation. A variety of ways to detect and compare these patterns are illustrated by the literature that will be summarized in this document and reviewed in detail in the Appendices. Pattern detection approaches include: visual analysis of patterns of metabolism; region of interest (ROI) analysis where the regions are hand-drawn or placed (sometimes with co-registration with anatomic images); and neural networks.

Kippenhan, et al. (1992), report that much of PET research involves improving the performance of particular links in the chain of highly complex data transformation that results in regional metabolic representations. Approaches to PET data management may include: normalization to a reference value (e.g., in brain studies to global brain metabolic rate or to an anatomic reference area that is relatively unaffected by the disease process) to generate metabolic ratios; or the use of absolute metabolic values. Links and Devous (1994) suggest that (for brain studies) the effect of normalization on diagnostic results may be dependent on region of interest (ROI) size, tomographic resolution, and biological and technical variation in the data and the type of normalization.

PET has been recognized as a valuable basic research tool during its approximately 20 years of development. Clinical diagnostic applications for PET are now emerging, particularly in the areas of neurology, cardiology, and oncology. Constructing and equipping, maintaining, and supporting PET facilities are resource-intensive activities requiring high levels of medical, technical, and managerial expertise. In the context of the widely recognized need to use available health care resources to maximize quality of care and achieve optimal patient outcomes, there is a compelling rationale for evaluating the clinical applications of PET as they emerge, and for applying evaluation results to policies regarding PET (Chalmers, 1988; Cooper, et al., 1988; Powers, et al., 1991; Hoffman, et al., 1992).

II. METHODS

A. Overview of the assessment methodology

The MDRC Technology Assessment Program convened a PET Advisory Committee, whose members are listed in *Appendix 1,* to focus the assessment. The Acting Under Secretary's question on experience with PET within VA was addressed by conducting surveys and site visits of VA PET centers to collect information on PET imaging utilization, center operations, and research activities. The results of this component of the assessment are outlined in *Appendix 9.*

PET has potential clinical applications in six conditions identified by the Advisory Committee as being of particular importance to the veteran population. These conditions are: solitary pulmonary nodules; lung cancer; head and neck cancers; breast cancer; colorectal cancer; and Alzheimer's disease. In the clinical management of these conditions, PET is applied as a diagnostic test.

One rationale for VA to invest in additional PET centers would be to make clinically useful PET studies more widely accessible to veterans. To respond to the Acting Under Secretary's question regarding whether to acquire additional PET capacity, a systematic review of research articles published in peer reviewed medical journals was used to evaluate what is known about the usefulness of PET in diagnosing diseases of importance to the veteran population. The systematic reviews *(Appendices 4 through 8)* addressed two additional queries:

> *Is PET an accurate diagnostic test when applied to patients with head and neck cancer, colorectal cancer, breast cancer, lung cancer/solitary pulmonary nodules, and Alzheimer's disease?*
>
> *Does PET affect patient management decisions, outcomes of care, costs of care, or cost-effectiveness of care in head and neck cancer, colorectal cancer, breast cancer, lung cancer/solitary pulmonary nodules, or Alzheimer's disease?*

The final literature database searches were performed on September 10, 1996; the assessment represents peer-reviewed literature published and indexed as of that date.

B. Systematic reviews in technology assessment

Expanding on the Banta and Luce (1993) systematic process for technology assessment, Goodman, et al. (1996) defined the following steps for conducting an assessment:

- identify assessment topics;
- specify the assessment problem;
- identify the locus for the assessment;
- retrieve evidence;
- collect new primary data (as necessary and appropriate);
- interpret the evidence;
- synthesize/consolidate the evidence;
- formulate findings and recommendations;
- disseminate findings and recommendations;
- monitor impact.

Banta and Luce (1993) note that synthesis is a critical part of the process. Synthesis involves a critical analysis of research results and other information, and often takes the form of judgments or recommendations. Synthesis is necessary to provide a responsible basis for decisions regarding the technology. Since policy makers are not generally trained in research study design and interpretation, raw data or unsynthesized results may be of little use to them. The purpose of synthesis is to make knowledge relevant to policy.

Synthesis provides focused, user-oriented information at a relatively low cost. If carefully performed, with attention to limitations of knowledge, synthesis can both guide technology-related decision making and help to define new research to answer important questions (Banta and Luce, 1993). However, the traditional narrative literature review has several shortcomings (Light and Pillemer, 1984; Mulrow, 1987 and 1994): the lack of formal rules for its conduct leads to subjectivity and bias; frequently used methods for synthesizing the results of multiple studies are inconsistent with good statistical practice; and it is an inefficient way to extract useful information.

Mulrow (1994) and other authors (e.g. Light and Pillemer, 1984; Slavin, 1986 and 1995) note that systematic reviews use a rigorous scientific approach and provide an alternative to traditional reviews. A systematic review frequently leads to different conclusions than does a traditional review of the same topic (Mulrow, 1994).

A systematic review or overview of the literature (Guyatt, et al., 1995):

- addresses a focused clinical question;
- uses appropriate criteria to select studies for inclusion;
- conducts a comprehensive search;
- appraises the validity of the individual studies in a reproducible fashion.

The purpose of a systematic review is to reduce unmanageable amounts of information to a form that is usable by decision makers, enabling health care decisions to be based on the best available evidence.

Systematic reviews can include both qualitative overviews of study findings and quantitative meta analyses of results. As currently understood, such reviews answer Slavin’s call (1986; 1995) to use “best evidence synthesis” to avoid the shortcomings associated with both qualitative reviews and indiscriminately applied meta analyses. A “best evidence” systematic review combines the quantification of effect sizes and systematic study selection procedures of quantitative syntheses with the attention to individual studies and methodological and substantive issues typical of the best narrative reviews. These reviews focus on the studies highest in internal and external validity, using well-specified and defended *a priori* inclusion criteria, and use effect size data as an adjunct to a full discussion of the literature being reviewed.

C. Search strategies

For each of the disease-specific systematic reviews conducted as part of this assessment of PET, literature was identified using formal search strategies. Comprehensive, multi-step search protocols were designed to ensure the broadest possible retrieval in each of the six disease areas: breast cancer, lung cancer, solitary pulmonary nodules, colorectal cancer, head and neck cancer, and Alzheimer's disease. Three searches for each disease were run on current files of the National Library of Medicine's MEDLINE®, and HEALTH® Planning databases for reviews of the literature, articles dealing with diagnosis, and for articles reporting the use of PET. Both free text words and MeSH subject headings were used to describe the concepts of interest and to ensure identification of the most comprehensive range of articles in the databases.

To ensure complete retrieval for the current period, when citations would not yet have appeared in the MEDLINE and HEALTH databases, searches were also performed on the ©Institute for Scientific Information's Current Contents® databases. Free text searches, using multiple synonyms, were employed for Current Contents searches.

Additional searches were performed for all of the cancers on the PDQ® Physicians' Data Query database (National Cancer Institute and National Library of Medicine). These non-bibliographic searches yielded information on diagnosis and staging of disease and on currently available treatment options; this information is incorporated into the background sections of the individual systematic reviews. All of the searches were refined according to the following rationale:

1. Early PET research in Alzheimer's disease used "first generation" scanners. These machines had limited spatial resolution, which contributed to potentially biased estimates of glucose metabolism due to partial volume effects (inclusion of cerebrospinal and subarachnoid spaces in the areas being analyzed for glucose metabolism) and the restriction of metabolic data to large neocortical areas. In addition, transmission scans were not used to correct for attenuation, venous blood was "arterialized" to estimate plasma radioactivity and glucose concentrations for metabolic rate calculations, and a highly subjective trace method was employed to determine regions of interest for analysis. Later generation scanners have improved resolution; the protocols used with these scanners correct for attenuation, collect arterial blood for metabolic calculations, and use devices to minimize patient movement during the relatively lengthy scanning procedures (Kumar, et al., 1991).

   The rapid evolution of the technology also affected the use of PET in oncology, and supported the restriction of the searches to the years 1991 to 1995. Significant articles appearing before that period were identified by selected searches of the years 1986 to 1991, and from the reference lists of the articles retrieved.

2. Publication in a peer-reviewed journal was required. This decision was based on preliminary review of the quality of studies in peer-reviewed journals, many of which failed to meet criteria for avoiding bias in diagnostic test evaluations. It was felt that abstracts that had not been subjected to the peer review process necessary for publication would have a high probability of representing studies of equivalent or lesser quality, and would generally also fail to meet criteria for diagnostic test evaluations.

3. Only studies using 2-[F-18]-2-deoxy-D-glucose (FDG) PET were included, as the majority of studies in oncology and Alzheimer's disease use this radiopharmaceutical.

D. Systematic review approach and protocol

The first question addressed by the diagnosis-specific systematic reviews of the literature concerned the accuracy of PET as a diagnostic test. The validity of the estimates of accuracy supplied by published studies was evaluated by applying a set of study design and reporting criteria from the methodologic literature, codified in the review protocol.

Accurate estimation of the characteristics of a diagnostic test is one of the early steps in the assessment of that test. However, accuracy does not extrapolate automatically to clinical utility, and a complete assessment requires further research. The second question (regarding outcomes of care) addressed by the reviews was focused by assigning published studies to levels in a hierarchy of "diagnostic efficacy", and by applying quality criteria (based on accepted principles of research design) appropriate to each level..

The diagnostic efficacy hierarchy explicitly acknowledges the goal of diagnostic testing to be improving processes of care, outcomes of care, and efficiency of resource use. It outlines the progression of research into a new diagnostic technology from the initial studies documenting technical performance of the imaging device, through accuracy studies, to studies documenting changes in treatment decisions based on diagnostic information, changes in outcome, and finally to studies of societal efficacy (i.e., cost-effectiveness, cost-benefit, or cost-utility studies from a societal perspective).

Each of the disease-specific systematic reviews was conducted using the following protocol, which codifies the analytic frameworks presented in *Appendix 2: Assessing Diagnostic Technologies.*

Systematic Review Protocol

1) Conduct MEDLINE and other database searches; retrieve full text articles that meet screening criteria:

- English language articles reporting primary data and published in a peer reviewed journal (not abstracts)
- studies≥ 12 human subjects (not animal studies) with the disease of interest (sample sized defined by PET Advisory Committee)
- studies using the radiopharmaceutical 2-[^{18}F]fluoro-2-D-glucose (FDG)

2) Apply screening criteria to bibliographies of retrieved articles as above, and retrieve additional articles.

3) Review full text articles and assign to level of Fryback and Thornbury (1991) diagnostic efficacy hierarchy.

4) Assign to **technical efficacy** level of Fryback and Thornbury diagnostic efficacy hierarchy:

- uncontrolled studies
- feasibility studies
- correlation studies of glucose metabolic rate changes with treatment

Systematic review protocol, continued

Studies whose stated purpose is to define diagnostic accuracy but which report results in a way that measures of diagnostic accuracy cannot be duplicated or interpreted, or in which some patients entered are not accounted for, will also be assigned to the technical efficacy level.

5) Assign to **diagnostic accuracy efficacy** level:

- stated purpose is to define diagnostic accuracy , and clinically useful measures (Se/Sp) provided or can be calculated
- meets full or modified (case series with internal controls; blinding if image analysis qualitative) evidence-based medicine criteria
- determines optimal cutpoint from ROC analysis or applies previously determined optimal cutpoint

Caveats will be attached to reports of sensitivity and specificity reported for case series with internal controls if prevalence of severe disease is high.

6) Assign to **diagnostic thinking efficacy** level if meets evidence-based medicine criteria (in box below) for evaluations of diagnostic tests and:

- numbers of subjects without target disorder ≥ numbers of cases with disorder (i.e., pretest probability of disease ≈ 50%)
- information useful in interpreting test results (i.e. converting pre- test probability of disease to post-test probability using predictive values or likelihood ratios) is provided or can be calculated from information in article.

Evidence-based medicine criteria for studies of diagnostic tests*

• Clearly identified comparison groups, ≥ 1 of which is free of the target disorder.

• Either an objective diagnostic standard (e.g. a machine-produced laboratory result) or a contemporary clinical diagnostic standard (e.g. a venogram for deep venous thrombosis) with demonstrably reproducible criteria for any subjectively interpreted component (e.g., report of better-than-chance agreement among interpreters).

• interpretation of the test without knowledge of the diagnostic standard result.

• Interpretation of the diagnostic standard without knowledge of the test result.

* Purpose and Procedure, *Evidence-Based Medicine*, November/December 1995

7) To further refine judgments of methodologic quality, grade **diagnostic accuracy or thinking efficacy** studies according to criteria in the box on the next page.

Systematic review protocol, continued

Methodologic quality of diagnostic accuracy and diagnostic thinking efficacy studies*

Grade	*Criteria*
A	**Studies with broad generalizability to a variety of patients and no significant flaws in research methods** • ≥ 35 patients with disease and ≥ 35 patients without disease (since such numbers yield 95% CIs whose lower bound excludes 0.90 if Se = 1) • patients drawn from a clinically relevant sample (not filtered to include only severe disease) whose clinical symptoms completely described • diagnoses defined by an appropriate reference standard • PET studies technically of high quality and evaluated independently of the reference diagnosis
B	**Studies with a narrower spectrum of generalizability, and with only a few flaws that are well described (and impact on conclusions can be assessed)** • ≥ 35 cases with and without disease • more limited spectrum of patients, typically reflecting referral bias of university centers (more severe illness) • free of other methods flaws that promote interaction between test result and disease determination • prospective study still required
C	**Studies with several methods flaws** • small sample sizes • incomplete reporting • retrospective studies of diagnostic accuracy
D	**Studies with multiple flaws in methods** • no credible reference standard for diagnosis • test result and determination of final diagnosis not independent • source of patient cohort could not be determined or was obviously influenced by the test result (work up bias) • opinions without substantiating data

* Adapted from:

Kent DL, Larson EB. Disease, level of impact, and quality of research methods: three dimensions of clinical efficacy assessment applied to magnetic resonance imaging. *IInvestigative Radiology* 1992; 27:245-54.

Kent DL, Haynor DR, Longstreth WT, Larson EB. The clinical efficacy of magnetic resonance imaging in neuroimaging. *Annals of Internal Medicine* 1994; 120:856-71.

8) Assign to **therapeutic efficacy level** if meets evidence-based criteria for evaluations of diagnostic tests and/or:

- authors discuss how test results did change, or could have changed, treatment for the patients enrolled in the study
- % of times subsequent procedure avoided due to test results, % of times prospectively stated therapeutic plans changed post-test documented.

9) Assign to **patient outcome efficacy** level if patient outcomes with PET are compared to those without PET in a case-control study, cohort study, or randomized controlled trial and/or:

- change in quality adjusted survival or cost/quality adjusted life year gained documented.

10) Assign to **societal efficacy** level if both costs (from a societal perspective) and consequences (efficacy, effectiveness, or utility) determined for both PET and an alternative.

11) Evaluate quality of studies at each efficacy level; conduct meta analyses if appropriate.

12) Articles are excluded from the review if they:

- are duplicated or superseded by subsequent study (at the same level of the hierarchy and with the same purpose) from the same institution
- contain insufficient information to judge comparability of case and control groups, details of imaging protocol, whether visual or quantitative analysis of PET data used, or type of PET quantitative data analysis used.

E. Meta analysis was considered for studies of PET diagnostic accuracy

Quantitative (statistical) pooling of the results of diagnostic accuracy studies to arrive at summary measures of sensitivity and specificity (defined in *Appendix 4*) or summary effect measures (Hasselblad and Hedges, 1995) for each disease was considered. The disease-specific systematic reviews described in *Appendices 3 through 8* resulted in decisions that meta analysis would not contribute to the assessment results; reasons for that decision are detailed in each appendix. In general, significant methodologic limitations that would tend to overestimate accuracy were present in all of the literature reviewed. These limitations argued against the validity and usefulness of pooling study results (Eysenck, 1994).

F. Selected alternatives to PET were addressed in the review

Brief discussions of alternate diagnostic technologies are included in the reviews. Alternate technologies were identified according to the following criteria:

- technologies that have been directly compared to PET;
- technologies that have been more rigorously assessed than PET (i.e. using stronger study designs and/or at a higher level in the diagnostic efficacy hierarchy) for a particular application, resulting in documentation that the alternate technologies are equally or more accurate and/or have a better defined role in patient management.

G. Review of the assessment report

The final draft of the assessment report was reviewed and approved by all members of the Advisory Committee, by one of the co-directors of the San Antonio Cochrane Center (Gilbert Ramirez, Ph.D.), and by the Under Secretary for Health, Kenneth W. Kizer, M.D., M.P.H. One of the committee members, an oncologist (Dr. Holohan), also reviewed all of the original literature cited in the oncology sections of the report to confirm the MDRC Technology Assessment Program’s evaluation of that literature. Other reviewers included: Alan Garber, M.D., Ph D. (methods), L. Jack Faling, M.D. and Charles Powell, M.D. (lung cancer, solitary pulmonary nodules), John Booss, M.D., Thomas Bird, M.D., Jeffrey Cummings, M.D., and Judith Salerno, M.D. (Alzheimer’s disease).

Summary of the Full Assessment Methodology

I. Systematic review of the literature
- Systematic review protocol applied.
- External reviewer independently judged studies for quality and position in hierarchy.
- Meta analysis was contemplated for diagnostic accuracy studies, but was not performed due to methodologic limitation in available studies.

II. Survey of VA PET facilities/site visits
- Data collected by written surveys covering fiscal years 1993, 1994, and 1995.
- Site visits conducted in August and September 1994
- Descriptive analysis and tabulation.

III. RESULTS: Site visits and surveys

PET is a relatively new addition to the repertoire of clinical diagnostic tests available both within and outside VA. Many of VA's PET facilities have become operational since 1990, and the information collected through the site visits and surveys represents preliminary data on VA experience with the technology.

The MDRC Technology Assessment Program obtained information on experience at eleven VA PET centers. A written survey and subsequent site visits were carried out from August through October, 1994. Another brief follow-up survey was distributed in December, 1995. Of the twelve initially approved PET sites, eleven were fully operational at the time of the assessment; support for the twelfth had been withdrawn. After completion of the site visits, support for another PET center was discontinued by the local VA medical center administration. At the time of release of this report, ten VA PET centers were in operation.

Interview subjects were selected by the VA PET centers, and included nuclear medicine physicians, PET center staff, referring physician specialists (in cardiology, neurology, oncology, and psychiatry, representing the clinical and research areas where PET is most commonly used), and hospital administrators. Most of the interview subjects had multiple job roles (administrative, clinical, and/or research), reflecting the academic environment for VA PET activity.

Full details of the site visit and survey findings are presented in *Appendix 9*.

A. Characteristics of sites

The pre-site visit surveys indicated that an equivalent array of ancillary services was offered at each location; substantial differences in PET utilization would not be likely to be attributable to differences in the types of patients (as represented by the range of services in place) treated at the VA medical centers (VAMCs). Most of the PET sites became operational in 1992 and 1993. The location of the PET camera was equally distributed between VAMCs and university affiliates, although the sharing partner/university affiliate tended to be the main source for radiopharmaceuticals. At most sites responsibility for PET center personnel was distributed equally between VA and university sharing partner.

Fluorodeoxyglucose (FDG) was the only radiopharmaceutical common to all sites. An important distinction among sites was the main mission or focus, which ranged from primarily research to primarily clinical work. The centers also differed in the models of PET scanners used.

B. Activity at each PET site

Activity at PET sites was compared using total number of patients studied rather than total number of scans. Expressing utilization according to the number of patients studied was felt to reflect most accurately the existing referral base for each site. While the total number of patients scanned is a relatively crude measure, adjusted measures reflecting variations among scanning protocols with respect to scan time and resources used, and PET technology across sites, would require a standardized workload unit and prospective data collection, neither of which has been systematically implemented across VA centers.

A wide range of types and volumes of PET studies was performed across VA in 1993 and 1994. During that time, more subjects appeared to have been scanned for clinical purposes than for research purposes. These trends may be attributed, in part, to differences in centers' definitions of studies classified as "clinical" versus those classified as "research."

- In 1993, most research activity system-wide was in neurology and psychiatry, followed by oncology; clinical activity consisted mainly of neurology and cardiology studies, followed by oncology.
- In 1994, most research activity was in neurology and psychiatry, followed by a growing interest in oncology; research activity in cardiology appeared to decrease. Neurology applications comprised the majority of clinical studies, followed by a growing interest in oncology applications. Some clinical cardiology studies were performed, but cardiology did not contribute substantially to overall clinical activity.
- In the brief follow-up survey conducted in 1995, six sites reported an increasing interest in clinical PET studies, which was attributed largely to increasing demand for clinical oncology studies. Interest in clinical cardiology applications continued to decrease. The increased interest in oncology studies may be attributed, in part, to the results of educational and marketing efforts made by PET center staff in recent years, and to the growing body of PET literature reporting clinical oncology applications. Two sites reported an increased use of PET in psychiatric and neurologic research.

C. Barriers and incentives to the use of PET

VA made a significant contribution to overall PET activity by committing substantial resources to the start-up of twelve PET centers. In return, PET has contributed significantly to overall research activity within VA. PET is regarded by many researchers in neurology and psychiatry as an essential tool for research into mental disorders, an area which is important to the veteran population. Additionally, many investigators view PET as a critical tool for basic physiologic research.

Foci of strong academic and clinical interests in functional imaging were important to obtaining initial support for PET at individual VA medical centers. Variations in current research activity across sites reflect the degree to which the initial interests extended into other research areas. The depth and breadth of the clinical and research bases at each site influenced the types of applications studied, the kinds of patients included in these studies, and the relative proportions of clinical and research studies conducted. At all sites, the reputation and expertise of the PET director and core PET center staff contributed to the willingness of medical staff and researchers to use PET as a clinical and research tool. The site visit interviews indicated that there are important organizational, professional, scientific, and reimbursement factors contributing to the relatively slow diffusion of PET into clinical practice. Interview subjects felt that limited FDA approved clinical PET applications and lack of demonstrated clinical utility perpetuated the perception of the general medical community and regulators that PET is primarily a research tool. Subjects felt that these factors also contributed to third party payers' inconsistent reimbursement policies.

PET is a very costly technology that requires a significant investment to cover start up costs and annual operating expenses. The major costs at each PET site were: equipment amortization; maintenance contracts for the scanner and cyclotron; scanner-related supplies; cyclotron supplies including target materials; and personnel, particularly highly skilled radiochemists, clinical and research specialists, analysts and programmers. Other significant costs included installation and maintenance of pneumatic tube systems used to transport radioactive isotopes between facilities.

PET directors and medical center directors have attempted to recover and reduce some of these costs. Those sites able to obtain reimbursement for clinical studies generally developed *a priori* consensus-building efforts among payers and providers within their communities in exchange for data collection. Multiple studies were often coordinated with production and use of radiotracers in an effort to minimize waste. Some sites generated revenue by selling cyclotron products, while others extended their catchment areas to include a broader patient base. One site made a decision to maintain low operating costs by purchasing cyclotron products from a private source, rather than producing its own. Two recommendations to offset the high and often unexpected maintenance costs of the scanner and cyclotron were made during the site visits: 1) establish an escrow account from equal contributions made by the sharing partners, and 2) support a "roving" maintenance team within VA to service all VA PET centers.

Inadequate staffing (particularly radiochemists) was cited as impeding the conduct of certain studies. Four PET centers cited the need for a qualified radiochemist as a major influence on the volume and variety of studies; competition for these specialists is intense. In VA hospitals, PET centers' hours of operation were frequently curtailed by inflexible tours of duty, restrictions in overtime salary, and restrictions and/or cutbacks in the number of Full Time Equivalent Employees. Reimbursement of patient transport costs for non-VA patients and the inability to transport less medically stable patients were barriers to access for some patients.

Competition among clinical specialties for access to PET, between PET and other technologies, and among PET centers in the same city may also affect access to PET for some patients. One center that developed a process to facilitate research protocol approval based on a NIH model; this assured equal representation of the sharing partners and the medical specialists interested in PET. Competition with other technologies and other local PET centers may dilute support for VA's PET facilities.

Several issues were related specifically to VA and to VA patients. Some interview subjects reported poor patient compliance in keeping scheduled appointments was noted. Others noted private sector patients' concerns about VA quality of care or perceptions that the services provided by PET centers at VA hospitals were restricted to VA patients only. Many VA PET center directors expressed frustration at not having the authority or resources to properly market their services to the private sector. The inability to attract VA patients for PET scans was attributed to either a lower burden of particular diseases among veterans compared to the general population, or to the failure of many veteran patients to meet protocol inclusion criteria.

Interview subjects saw centralized strategic planning around distribution, construction and maintenance as necessary to the overall investment in costly technologies such as PET. Nevertheless, subjects described these processes as frustrating, inefficient, and protracted. Local VA administrators perceived a lack of vision and commitment to PET by Headquarters; many felt that they were expected to support new, costly programs and services within existing funding levels.

Variations in VA's financial commitment among the centers appeared to be related to the degree to which local medical center directors sustained the support, often through the sharing agreements with academic affiliates. The agreement negotiating team typically included representatives from Fiscal Service and the Director's Office. The degree to which the Director's Office participated in these negotiations varied across sites; the most active participation tended to produce some of the most functional arrangements. To comply with VA policy, PET center directors with dual appointments were excluded from negotiations. Consequently, interview subjects felt that the negotiations could not benefit from the insight of the individual who was most familiar with the needs of the center.

In these agreements, PET center cost sharing varied; under some agreements, costs were evenly distributed between partners, while other agreements stipulated alternate means of distributing costs. VA's contribution ranged from covering partial costs of the scanner to covering partial costs of both the scanner and overhead. Unrealistically high volume projections and unreasonably low overhead costs formed the basis on which some of the original sharing agreements were negotiated. Negotiations in recent years have used more realistic volume projections or a patient charge based on the national average. One site developed a workload unit to better reflect true utilization of resources.

Interview subjects felt that the sharing arrangement most favorable for VAMCs with PET centers located at the academic affiliate was one that required full payment up front by the affiliate for its portion of the scanner. If contributing to overhead costs, the VAMC was subsequently billed on a fee-for-service basis at a charge approximately equal to the national average. Another arrangement favorable to the VAMC was one in which a fixed number of "free" scans for VA patients was determined up front, in exchange for partial use of the scanner by other sharing partners. These arrangements insure that each VAMC recovers its portion of the investment up front, without risk of financial loss, should volume projections be unfulfilled or overhead costs be excessive.

IV. RESULTS: Systematic reviews

The full background, results, and discussion texts, data abstraction tables for diagnostic accuracy and therapeutic efficacy studies, and comparisons of PET to alternate technologies are presented in the appendices (*Appendix 3: Head and Neck Cancer; Appendix 4: Colorectal Cancer; Appendix 5: Breast Cancer; Appendix 6: Lung Cancer; Appendix 7: Solitary Pulmonary Nodules; and Appendix 8: Alzheimer's Disease).* The overall results of the systematic reviews are summarized here in Tables 1 through 6 (pages 20 through 28).

The systematic reviews indicate that research into the clinical utility of PET in selected conditions relevant to the veteran population is in its preliminary stages. The available studies have focused on the feasibility of using PET in these conditions, and on defining its accuracy as a diagnostic test. A few studies have addressed changes in treatment decisions based on PET findings. However, the MDRC was unable to locate any studies documenting changes in outcomes of care or costs of care associated with incorporating PET into diagnostic strategies for the conditions addressed in this assessment. Critical research into defining the clinical consequences of using PET for diagnosis has yet to be performed or reported.

Since most of the PET studies analyzed for the systematic reviews address diagnostic accuracy, revisiting criteria for a valid evaluation of diagnostic test accuracy is advisable here. The McMaster University Department of Clinical Epidemiology and Biostatistics provided a seminal (1981) and concise list of the questions to ask regarding published clinical evaluations of diagnostic tests. These are:

- Was there an independent, "blind" comparison with a "gold standard" of diagnosis?
- Did the patient sample include an appropriate spectrum of mild and severe, treated and untreated disease, plus individuals with different but commonly confused disorders?
- Was the setting for the study, as well as the filter through which study patients passed, adequately described?

- Was the reproducibility of the test result (precision) and its interpretation (observer variation) determined?

- Was the term “normal” defined sensibly?

- If the test is advocated as part of a cluster or sequence of tests, was its contribution to the overall validity of the cluster or sequence determined?

- Were the tactics for carrying out the test described in sufficient detail to permit their exact replication?

- Was the “utility “ of the test determined?

According to these and analogous criteria incorporated into this assessment’s systematic review protocol, the published studies using PET to diagnose Alzheimer’s disease have been relatively well constructed and present a coherent set of observations on PET’s good level of agreement with widely used clinical criteria for dementia of the Alzheimer’s type. As most results fell within a relatively narrow range of estimates of accuracy, meta analyses of the diagnostic accuracy results were not conducted.

There are, however, barriers to moving PET into routine use in diagnosing Alzheimer’s disease and to affecting outcomes of care by means of PET diagnosis:

- relatively few of the published studies prospectively evaluated large numbers of patients with causes of dementia that can be confused, or present concurrently, with Alzheimer’s disease, making a valid estimate of the positive predictive value of PET difficult to determine;

- histologically verified Alzheimer’s disease represents a subset of patients with clinically diagnosed dementia of the Alzheimer’s type, and the results of ongoing studies defining the agreement of PET with the gold standard of autopsy diagnosis in Alzheimer’s disease are not yet available;

- effective treatments for Alzheimer’s disease are not available. An accurate diagnostic test (relative to the gold standard of autopsy result) is needed for research into treatments for Alzheimer’s disease; both PET and other tests that have been shown to have equivalent accuracy may be useful in this role.

The published evidence for the accuracy of PET in diagnosing cancer is less convincing than that for its accuracy in diagnosing dementia of the Alzheimer’s type. While the available studies report good face accuracy for PET (particularly in clinical settings where PET was used to differentiate recurrent cancer from treatment artifacts such as scars), many of the studies did not adhere to the principles of study design outlined above.

Almost all PET cancer studies are retrospectively analyzed case series. They enrolled relatively few patients (too few to allow one to comfortably draw conclusions from the data), did not include control groups (i.e., did not adequately account for biologic variation in test results or differential diagnosis with other conditions), and, when PET images were visually interpreted, often did not blind image interpreters or address issues of interobserver variation. Many published studies can be assumed to be subject to context bias (Egglin and Feinstein, 1996). The studies that compared PET to other diagnostic technologies did not randomize the order of test administration, and in some cases were subject to work up bias (where the results of one test led to the decision to perform another, or to confirm diagnosis by biopsy); these biases will have affected accuracy estimates of both PET and the alternative test or tests.

A critical shortcoming in the diagnostic accuracy PET oncology literature for organizations, like VA, that are seeking to rationalize the provision of services on a regional or system-wide basis, is the lack of epidemiologic information in the published studies. The filters through which patients passed to be included in the published case series are often inadequately described, making extrapolation of the results to defined populations, and subsequent planning for these populations, difficult. To summarize, like the early studies into other diagnostic technologies, the methodologic weaknesses of the available PET oncology studies will have tended to overestimate accuracy and clinical value. Accordingly, meta analyses of these studies were not performed.

A widely credited role for PET in the nuclear medicine and surgery literature is that of increasing diagnostic certainty regarding the need for invasive procedures (e.g., neck dissection in patients with head and neck cancer, resection of metastases from colorectal cancer that are potentially curable if isolated, axillary dissection in breast cancer, thoracotomy for solitary pulmonary nodules). The authors of a few oncology studies discussed the potential or actual changes in treatment that resulted from incorporating PET into diagnostic strategies at critical decision points in cancer treatment processes. These studies were retrospective case series that had not been specifically designed to document changes in treatment; methods for recording changes in treatment plans were not specified, and results data tended not to be systematically analyzed or presented. The studies generally enrolled highly selected patients whose previous work-up was not clearly specified, nor was the size or composition of the referral base from which the patient sample was drawn. Information from PET studies resulted in more appropriate treatment for some patients. However, the published studies tended to give inadequate details about what happened to patients whose PET studies did not accurately reflect their disease status.

PET is generally presented as complementary to anatomic imaging studies such as CT or MRI. Accordingly, further work on PET's treatment impact and role in a multi-test diagnostic strategy is needed before the population impact of PET on a health care system such as VHA can be estimated. It should be noted that the management of patients with cancer is a highly complex area, with many uncertainties beyond those related to the impact of PET; specifically, treatment for many cancers (particularly the solid tumors addressed in this assessment) is less than optimally effective. Before population outcomes (e.g., mortality rates from specific cancers) can be improved, a wide range of interventions for prevention, screening, diagnosis, treatment, and follow-up also need to be improved.

The disease-specific systematic reviews in this assessment included, for comparative purposes, information on some of the diagnostic technologies that may be alternatives to PET. While information on these alternatives was not identified and retrieved with the same thoroughness as the information on PET, a sample of articles from the recent peer reviewed literature indicates that research into alternate tests has resulted in substantial improvements in accuracy for many of the conditions discussed in this assessment. Some of the research that has been conducted for alternative diagnostic tests surpasses the PET literature in its methodologic rigor.

Systematic review summary Tables 1 through 6 (pages 20 through 28) present the results of diagnostic accuracy efficacy studies in each of the diseases considered for this assessment. Studies were included if they met all or some of the evidence-based medicine criteria for diagnostic test evaluations. Methodologic quality grades, which further refine the quality judgments implicit in the evidence-based medicine criteria, are also noted. Many of the included studies do not meet high methodologic standards; in the absence of more rigorous studies, they are presented here in an effort to make the review methods and conclusions as transparent as possible.

V. RISKS ASSOCIATED WITH PET RADIOPHARMACEUTICALS

Research articles reviewed for this assessment either:

- provided no comments on any risks associated with PET radiopharmaceuticals, or
- included a statement indicating that no patients in the study experienced adverse events after radiopharmaceutical administration.

VI. FDA STATUS OF PET RADIOPHARMACEUTICALS

The Food and Drug Administration has approved two PET radiopharmaceuticals. The quotations below are from the package inserts.

- CardioGen (Rubidium Rb 82 Generator) is indicated for use as “a myocardial imaging agent that is useful in distinguishing normal from abnormal myocardium in patients with suspected myocardial infarction.”

- Fludeoxyglucose F 18 Injection [(^{18}F-FDG) The Methodist Medical Center of Illinois] is indicated for “the identification of regions of abnormal glucose metabolism associated with foci of epileptic seizure. FDG is not indicated for distinguishing epileptogenic foci from brain tumors or other brain lesions which may cause seizures.”

The many PET imaging facilities that “compound” radiopharmaceuticals on-site do so under the aegis of state practice of medicine and pharmacy laws. The FDA has determined that PET centers are manufacturing a new drug and that they are subject to existing new drug regulations. Facilities that manufacture radiopharmaceuticals for clinical use must file a New Drug Application (NDA) or an Abbreviated New Drug Applications (ANDA) and conform with Current Good Manufacturing Practice (CGMP) standards. Clinical investigators who wish to conduct clinical trials with an unapproved PET radiopharmaceutical must file an Investigational New Drug (IND) application for each drug.

The FDA has offered to work with the PET community to help sponsors and investigators interpret and utilize the appropriate regulations in offer to comply with the existing new drug regulations.

VII. CONCLUSIONS

The site visits and surveys confirm that VA has made a substantial resource commitment to its PET imaging facilities. This commitment has the potential to contribute to fulfilling two parts of VA's mission: research and clinical care. PET is widely credited as an important basic research tool in the literature; VA PET researchers who were interviewed for this assessment share this belief. The efficiency of basic research activities would be enhanced by implementation of suggestions for improving operations that were made by VA PET center personnel during the site visits.

The site visits and surveys outlined a wide range of research and clinical activities in VA PET centers. There are many site-specific protocols and areas of research interest. Coordination of these activities has not been systematically addressed.

The presence of PET on the lists of many health care assessment agencies (*Appendix 10*), nationally and internationally, and many discussions in the medical literature attest to concerns that PET will follow a familiar diffusion trajectory into clinical care before its usefulness and contribution to improved outcomes have been adequately evaluated. The trends seen in PET utilization during the assessment period indicate that oncology is an increasing focus for clinical activity in VA. This trend should be of interest to policy makers in the context of the findings of the systematic reviews reported here.

This assessment's systematic reviews of the literature indicate, to the extent that the published literature represents the existing data, that the knowledge base supporting clinical diagnostic applications of PET has significant deficiencies. Methodologic weaknesses in published studies seriously limit the validity and generalizablity of the available evidence on the accuracy of PET as a diagnostic test, and PET's contribution to improving outcomes has not been systematically addressed. Accordingly, the assessment team believes that the literature as of September, 1996 does not support widespread incorporation of PET studies into routine diagnostic strategies for the applications addressed in this assessment.

The Advisory Committee to the PET assessment believed that the assessment results supported a conclusion that VA should maximize the value derived from its existing resource commitment, rather than invest in additional PET centers at the present time. Maximizing the value of the existing commitment could include:

- Building organizational structures to coordinate its PET activities across the VA system.
- Implementing a VA PET registry. Systematic, standardized data (including those related to work load, resources used, and operations) specific to PET would facilitate future assessment efforts. A registry for tracking diagnosis-specific utilization, the marginal contribution of PET to a diagnostic strategy involving other tests, impact of PET on treatment decision making, and treatment outcomes would also facilitate future assessments.
- Organizing a cooperative group consisting of VA PET centers and their academic affiliates. Such a group could facilitate efforts to comply with FDA regulations. Efficacy research in oncology is frequently conducted by cooperative groups, and could supply a model for PET oncology research. A VA cooperative PET group could also attempt to define clinical research areas of interest to the entire VA system, and to design multi-center studies of high methodologic quality.
- Supporting rigorous, prospectively designed clinical research that corrects the methodologic limitations outlined in the diagnosis-specific systematic reviews. Once the diagnostic accuracy of PET has been adequately defined, attention should be directed to defining the changes in patient management decisions, outcomes of care, patient outcomes, cost-

effectiveness of care, and cost-utility of care that are associated with incorporating PET into diagnostic strategies.

- Submitting currently unpublished data from studies of high methodologic quality for peer review. Advocates of PET both within and outside VA feel strongly that the clinical utility of PET is increasingly evident; these opinions may be supported by currently unpublished data.

Table 1 Summary of the literature: Diagnostic accuracy efficacy of PET and alternatives in head and neck cancer
(from studies comparing PET directly to other diagnostic tests)

Role	Study	N	Operating characteristics*				Evidence-based medicine criteria**			Methodologic quality grade***
			PET	*CT*	*MRI*	*Other*	*controls*** **	*standard*	*blinding*	
Unknown primary	Rege, et al., 1994	4 cases 0 controls	Se = 50%		Se = 0%		-	+	-	D
Known primary site	Rege, et al., 1994	30 cases 0 controls	Se = 97%		Se = 77%		-	+	-	D
	Laubenbacher, et al., 1995	17 cases 0 controls	Se = 100%		Se = 100%	endoscopy, Se = 100%	-	+	-	D
Primary tumor staging (size, extent)	Laubenbacher, et al., 1995	17 cases 0 controls	Se = 41%		Se = 41%	endoscopy, Se = 59%	-	+	-	D
Cervical node involvement	Rege, et al., 1994	16 pos 18 neg	Se = 88% Sp = 89%		Se = 81% Sp = 89%		+	+	-	D
	McGuirt, et al., 1995	14 pos 31 neg	accuracy = 82%	accuracy = 82%		clinical exam accuracy = 71%	+	+	-	D
	Laubenbacher, et al., 1995	83 pos nodes 438 neg nodes	Se = 90% Sp = 96%		Se = 78% Sp = 71%		+	+	-	D
		18 pos neck sides 16 neg neck sides	Se = 89% Sp = 100%		Se = 72% Sp = 56%		+	+	-	D
	Braams, et al., 1995	22 pos nodes 177 neg nodes	Se = 91% Sp = 88%		Se = 36% Sp = 94%		+	+	-	D
	Benchaou, et al., 1996	54 pos node groups 414 neg node groups	Se = 72% Sp = 99% PPV = 89% NPV = 99%	Se = 67% Sp = 97% PPV = 74% NPV = 95%		clinical exam Se = 61% Sp = 97% PPV = 72% NPV = 95%	+	+	+	B
Suspected recurrent disease	Rege, et al., 1994	10 pos 7 neg	Se = 90% Sp = 100%		Se = 67% Sp = 57%		+	+	-	D
	Lapela, et al., 1995	16 pos 17 neg	Se = 88 -94% Sp = 43 -86% depending on criteria for pos	Se = 92% Sp = 50%			+	+	+	C

Abbreviations:
Ct, computed tomography
MRI, magnetic resonance imaging
neg, negative for disease
pos, positive for disease
Se, sensitivity
Sp, specificity
PPV, positive predictive value
NPV, negative predictive value
US/FNA, ultrasound/fine needle aspiration

* operating characteristics defined in Appendix 2: Assessing Diagnostic Technologies, pages 5-7
** Appendix 2, page 8
*** Appendix 2, page 9

****"controls" were case series patients with benign conditions

Table 2 **Summary of the literature: Diagnostic accuracy efficacy of PET and alternatives in colorectal cancer**

Notes *The PET studies in this table were retrospectively analyzed case series; internal controls (cases with benign, rather than malignant, conditions) allowed the calculation of specificity as well as sensitivity. Some of the alternatives to PET have been evaluated using more rigorous study designs.*

Some studies analyzed results separately according to the clinical role of PET for subsets of patients; these studies appear in the table more than once, and may have received different methodologic quality grades for each subset analysis.

Role	Study	N	Operating characteristics*				Evidence-based medicine criteria**			Methodologic quality grade***
			PET	***CT***	***MRI***	***Other***	***controls***	***standard***	***blinding***	
Detecting or staging primary or recurrent disease	Falk, et al., 1994	16 patients: 15 malignant lesions; 3 benign lesions	Se = 87% Sp = 67%	Se = 47% Sp = 100%			+ (internal)	+	partial	D
	Nattinger, et al., 1991 (ACP review)					colonoscopy Se = 94% Sp = 100%	(review)	(review)	(review)	(review)
	Hernandez-Socorro, et al., 1995	40 cases 64 controls				colonoscopy Se = 94% Sp = 100% hydrocolonic ultrasound Se = 97% Sp = 97%	+	+	+	B
Diagnosing recurrent tumor vs scar	Strauss, et al., 1989	29 patients: 21 malignant lesions; 8 scar	Se = 95% Sp = 100%				+ (internal)	+	+ (quantitative analysis)	C
	Schlag, et al., 1989	18 patients: 11 malignant lesions; 6 scar	Se = 92% Sp = 100%			immunoscintigraphy Se = 40% Sp = 50%	+ (internal)	+	+ (quantitative analysis)	C
	Ito, et al., 1992	15 patients: 11 malignant lesions; 4 scar	Se = 100% Sp = 100%		Se = 91% Sp = 100%		+ (internal)	+	+ (quantitative analysis)	C
	Schiepers, et al., 1994	6 patients: 5 malignant lesions; 1 scar	Se = 100% Sp = 100%				+ (internal)	+	-	D
Diagnosing recurrent tumor vs scar	Hawes, et al., 1993	85 with disease 408 without disease (review with weighted average of results from 7 studies)				endoscopic ultrasound Se = 99% Sp = 88%	(review)	(review)	(review)	(review)

Role	Study	N	Operating characteristics*				Evidence-based medicine criteria**			Methodologic quality grade***
			PET	*CT*	*MRI*	*Other*	*controls*	*standard*	*blinding*	
Diagnosing liver metastases	Schiepers, et al., 1994	80 studies: 34 malignant lesions; 46 benign lesions	Se = 94% Sp = 100%			CT and/or ultrasound Se = 85% Sp = 98%	+	+	-	C
	Vitola, et al., 1996	55 sites: 39 malignant; 16 benign 24 patients: 19 malignant disease; 5 benign	Se = 90% Sp = 100% Se = 95% Sp = 100%	Se = 86% Sp = 58%		CT portography Se = 97% Sp = 9% Se = 100% Sp = 33%	+ (internal)	+	+ (semiquantitative analysis)	C
	Lai, et al., 1996	34 patients: 27 with malignant disease; 7 benign or no disease	Se = 93% Sp = 57%	Se = 100% Sp = 14%	Se = 100% Sp = 80%		+ (internal)	+	+	C
	Stark, et al., 1987	57 cases; 72 controls: 21 benign liver disease; 51 with normal livers		Se = 80% Sp = 94%	Se = 82% Sp = 99%		+	+	+	B
	Panzer, et al., 1991 (ACP review)	review		Se = 90% Sp = 90% LR + = 8 LR - = 0.11		ultrasound, adequate studies Se = 80% Sp = 90% LR + = 9 LR - = 0.22	(review)	(review)	(review)	(review)
Diagnosing liver metastases	Rafaelsen, et al., 1995	295 patients: 64 with liver metastases 231 without liver metastases				liver enzymes Se = 9-47% Sp = 92-98% preop US Se = 70% Sp = 94% surgical exploration Se = 84% Sp = 97% intraop US Se = 97% Sp = 98%	+ (internal)	+	+	B

Abbreviations
CT, computed tomography
MRI, magnetic resonance imaging
neg, negative for disease
pos, positive for disease
LR, likelihood ratio
PPV, positive predictive value
NPV, negative predictive value
US/FNA, ultrasound/fine needle aspiration
ACP, American College of Physicians

*operating characteristics defined in Appendix 2: Assessing Diagnostic Technologies, pages 5-7
**Appendix 2, page 8
*** Appendix 2, page 9

Table 3 Summary of the Literature: Diagnostic accuracy efficacy studies of PET and alternatives in breast cancer

Notes: All studies except Nieweg, et al., 1993b, which was a case-control study, were series of patients presenting for surgical evaluation of breast masses (a high index of suspicion of malignant disease) and included internal controls as the comparison group. Predictive values should be viewed accordingly. Studies assessing axillary node involvement included patients with malignant primary breast disease. Results from Avril, et al., 1996b were reported as ranges of data from all subgroup analyses. Results from Avril, et al., 1996a included all patients with benign and malignant primary disease and represent 95% confidence intervals; subgroup analyses were not reported because of their small study size. None of these studies met strict evidence-based medicine criteria for blinding, but all studies provided data on the comprehensiveness of blinding of test interpreters to the gold standard.

Where substantial duplication in purpose of study, patients studied, and results in multiple studies from the same institution could be inferred, only the most recent, largest, most rigorously designed, or most comprehensive was included in the table. Although data from both studies by Avril and associates (1996a and 1996b) represent the same patient population, these studies addressed different purposes; inclusion of both publications were felt to be warranted.

Abbreviations are listed at the end of the table.

Role	**Study**	**N**	**Operating Characteristics***			**Evidence-Based Medicine Criteria****			**Methodologic Quality Grade*****
(Note: some studies assessed multiple roled)			***PET***	***Clinical Exam***	***Mammography***	***comparison group***	***histologic gold standard***	***blinding***	
Defining primary disease	Adler, et al., 1993	27 positive lesions 8 negative lesions	Se=96% Sp=100%			+ internal	+	+	**C**
	Nieweg, et al., 1993b	11 cases 8 controls	Se=91% Sp=100%			+	+	+	**C**
	Avril, et al., 1996b	41 positive lesions 31 negative lesions	Se=68%-94% Sp=84%-100% PPV=87%-97% NPV=70%-93%			+ internal	+	partial	**D**
	Scheidhauer, et al., 1996	23 malignant cases 7 benign cases	Se=91% Sp=86%	Se=74% Sp=71%	Se=86%	+ internal	+	partial	**D**
Defining axillary node involvement	Adler, et al., 1993	9 positive axillae 10 negative axillae	Se=90% Sp=100%			+ internal	+	+	**C**
	Avril, et al., 1996a	24 positive axillae 27 negative axillae	Se=57%-93% Sp=81%-100% PPV=75%-100% NPV=66%-100%	Se=36%-78% Sp=66%-96% PPV=30%-70% NPV=51%-85%		+ internal	+	+	**C**
	Scheidhauer, et al., 1996	9 malignant cases 9 benign cases	Se=100% Sp=89%			+ internal	+	partial	**D**
Detecting distant metastases	Scheidhauer, et al., 1996	8 positive lesions 15 negative lesions	Se=100% Sp=100%			+ internal	+	partial	**D**

N, number of study subjects included in analysis; unless otherwise noted, data are analyzed by subject
Se, sensitivity
Sp, specificity
PPV, positive predictive value
NPV, negative predictive value

* operating characteristics defined in *Appendix 2: Assessing Diagnostic Technologies, pages 5-7*
** *Appendix 2, page 8*
*** *Appendix 2, page 9*

Table 4 Summary of the Literature: Diagnostic accuracy efficacy studies of PET and alternatives in lung cancer

Notes: All of the studies in the table are case series (Level V evidence) with internal controls (i.e. those with benign masses) used as a comparison group. All patients in these studies had suspected or biopsy-proven lung cancer (i.e. the pre-test probability of disease in the study populations was very high). Results from Knight, et al., 1996 and Inoue, et al., 1995 were reported as ranges to include data from all subgroup analyses.

None of these studies met strict evidence-based medicine criteria for blinding, but all studies presented information on blinding of the test interpreters to the biopsy gold standard. Blinding of the PET interpreters to other clinical and radiologic data varied across studies and is reflected in the columns designated "Operating Characteristics"; "PET + CT" indicates a complementary role of PET with CT, and PET alone indicates a substitutive role of PET for CT.

Where substantial duplication in purpose of study, patients studied, and results in multiple studies from the same institution could be inferred, only the most recent, largest, most rigorously designed, or most comprehensive was included in the table. Studies reviewed but not included are listed under "References".

Abbreviations are listed at the end of the table.

Role	***Study***	***N***	***Operating Characteristics****			***Evidence-Based Medicine Criteria*****			***Methodologic Quality Grade******
(Note: Some studies assessed multiple roles)			***PET***	***PET + CT***	***CT***	***comparison group***	***histologic gold standard***	***blinding***	
Defining unknown primary disease	Kubota, et al., 1990	12 malignant cases 10 benign cases		Se=83% Sp=90% accuracy=86%	no data reported	+ internal	+	+	**C**
	Scott, et al., 1994	47 malignant cases 15 benign cases		Se=94% Sp=80%	no data reported	+ internal	+	+	**C**
	Slosman, et al., 1994	31 malignant cases 5 benign cases		Se=93.5%	no data reported	+ internal	+ & follow-up	+	**C**
	Wahl, et al., 1994	19 malignant cases 4 benign cases	Se=100%		Se=100%	+ internal	+	+	**C**
	Sazon, et al., 1996	82 malignant cases 25 benign cases	Se=100% Sp=52%		no data reported	+ internal	+	+	**C**
	Knight, et al.,1996	32 malignant cases 16 benign cases		Se=100% Sp=58%-63% PPV=75% NPV=100%	Se=33%-41% Sp=52% PPV=83% NPV=52%	+ internal	+	+	**D**

Role (Note: Some studies assessed multiple roles)	Study	N	Operating Characteristics*			Evidence-Based Medicine Criteria**			Methodologic Quality Grade***
			PET	PET + CT	CT	comparison group	histologic gold standard	blinding	
Detecting overall lymph adenopathy	Patz, et al., 1995	42 patients with: 23 malignant nodes 39 benign nodes	Se=83% Sp=82%		Se=43% Sp=85%	+ internal	+	+	D
Detecting hilar/lobar lymph adenopathy	Patz, et al., 1995	42 patients with : 11 malignant nodes 29 benign nodes	Se=73% Sp=76%		Se=27% Sp=86%	+ internal	+	+	D
Detecting mediastinal lymph adenopathy	Patz, et al., 1995	42 patients with: 12 malignant nodes 10 benign nodes	Se=92% Sp=100%		Se=58% Sp=80%	+ internal	+	+	D
	Wahl, et al., 1994	23 patients with: 11 malignant sides 16 benign sides	Se=82% Sp=81% accuracy=81%		Se=64% Sp=44% accuracy=52%	+ internal	+	+	C
	Chin, et al., 1995	9 malignant cases 21 benign cases		Se=70% Sp=81% accuracy=80%	Se=56% Sp=86% accuracy=77%	+ internal	+	+	D
	Valk, et al., 1995	24 malignant sides 52 benign sides		Se=83% Sp=94% accuracy=91%	Se=63% Sp=73% accuracy=70%	+ internal	+ & follow-up	+	D
	Sazon, et al., 1996	32 patients with: 16 malignant sides 16 benign sides	Se=100% Sp=100%		Se=81% Sp=56%	+ internal	+	+	C
	Scott, et al., 1996	10 malignant nodes 65 negative nodes within: 9 malignant cases 18 benign cases		Se=100% Sp=98%-100%	Se=60% Sp=83%-94%	+ internal	+	+	D
Distinguishing local cancer recurrence from fibrosis	Patz, et al., 1994	35 recurrence cases 8 fibrosis cases		Se=97.1% Sp=100%	no data reported	+ internal	+ & follow-up	+	D
	Inoue, et al., 1995	23 recurrence cases 13 fibrosis cases		PET + x-ray, CT, MRI Se=100% Sp=56%-78% accuracy=86%	no x-ray, CT, or MRI data reported	+ internal	+ & follow-up	+	D

N, number of total study subjects included in analysis; unless otherwise noted, data are analyzed by subject
Se, sensitivity
Sp, specificity
PPV, positive predictive value
NPV, negative predictive value
CT, computed tomography
MRI, magnetic resonance imaging

** operating characteristics defined in Appendix 2: Assessing Diagnostic Technologies, Pages 5-7*
*** Appendix 2, page 8*
**** Appendix 2, page 9*

Table 5 Summary of the Literature: Diagnostic accuracy efficacy studies of PET in solitary pulmonary nodules

Notes: All of the studies in the table are case series (Level V evidence) and met most of the evidence-based medicine criteria for diagnostic test evaluations. None of the studies met strict evidence-based medicine criteria for blinding, but all studies except Gupta, et al., 1996 provided information on the comprehensiveness of blinding of test interpreters to the biopsy gold standard. Blinding of PET interpreters to other clinical and radiologic data varied across studies.

Internal controls (i.e. those with benign masses) were used in each study, and it was possible to calculate sensitivity and specificity for PET in those studies. The pre-test probability of disease in these study populations was very high, and predictive values were not reported. Each study varied in inclusion criteria with respect to maximal lesion size and image characteristics (pulmonary masses, ill-defined infiltrates, focal lesions). Operating characteristics from these studies should be interpreted with caution.

Where substantial duplication in purpose of study, patients studied, and results in multiple studies from the same institution could be inferred, only the most recent, largest, most rigorously designed, or most comprehensive was included in the table. While data from Dewan, et al., 1995 and Gupta, et al., 1996 are likely derived from the same patient population, these studies addressed different purposes, and inclusion of both was felt to be warranted. Studies reviewed but not included are listed under "References".

Abbreviations are listed at the end of the table.

Role	***Study***	***N***	***Operating Characteristics****			***Evidence-Based Medicine Criteria*****			***Methodologic Quality Grade*****
			PET	***TTNA***	***other***	***comparison group***	***histologic gold standard***	***blinding***	
Defining unknown SPN	Dewan, et al., 1995	26 malignant lesions 9 benign lesions	Se=100% Sp=78% accuracy=94%	Se=81% Sp=100% accuracy=86%		internal	+	partial	**D**
	Bury, et al., 1996	33 malignant cases 17 benign cases	Se=100% Sp=88%			internal	+	+	**C**
	Duhaylongsod, et al., 1995b	59 malignant cases 28 benign cases	*for lesions < 4 cm* Se=97% Sp=81% accuracy=92%			internal	+	+	**C**
	Gupta, et al., 1996	45 malignant cases 16 benign cases	Se=93% Sp=88% accuracy=92%			internal	+	unclear	**C**

Se, sensitivity
Sp, specificity
TTNA, transthoracic needle aspiration biopsy

** operating characteristics defined in Appendix 2: Assessing Diagnostic Technologies, pages 5-7*
*** Appendix 2, page 8*
**** Appendix 2, page 9*

Table 6 Diagnostic accuracy and diagnostic thinking efficacy of PET and its neuroimaging alternatives

Notes: *All of the studies that evaluated a diagnostic test against the standard of histopathology fully met evidence based medicine and other methodologic quality criteria (i.e., received methodology grades of A or B). PET and other studies evaluating the new technology against clinical criteria for dementia of the Alzheimer s type would receive methodology grades of A or B, with the exception of the absence of histopathologic diagnosis.*

Neuroimaging Test	Diagnostic Standard Used in Evaluation Studies		Characteristics
	Histopathology	*Clinical criteria*	
CT	x		Se = 94%; Sp = 93.5% (AD-specific orientation; AD vs normal controls and other dementias) *Jobst, et al., 1994*
SPECT	x		Se = 96%; Sp = 89% (AD vs normal controls and other dementias) *Jobst, et al., 1994*
		x	Sp = 89% • all probable AD, Se = 43% • probable AD < 80 years, Se = 56% • probable AD > 80 years, Se = 29% • SPECT contributed to 8% of final diagnoses *Van Gool, et al., 1995*
CT + SPECT	x		Se = 90%; Sp = 97% (AD vs normal controls and other dementias) *Jobst, et al., 1994*
PET		x	Se = 94.6; Sp = 97% ("robust ratio"; DAT vs normal controls) *Herholz, et al., 1993* Post test probability of disease, positive test = 90%; posttest probability, negative test = 10% in patients with pretest probability of disease = 50% (neural net; DAT vs normal controls) *Kippenhan, et al., 1994* Se = 94%; Sp = 79% (4 image patterns typical of DAT; DAT vs normal controls) *Salmon, et al., 1994* Se = 94%; Sp = 53% (4 image patterns typical of DAT; DAT vs non-DAT dementia controls) *Salmon, et al., 1994* Se = 94%; Sp = 99% (stereotactic surface projections; DAT vs non-DAT controls) *Burdette, et al., 1996*
PET vs CT		x	PET: Se = 97%; Sp = 84% (qualitative) CT: Se = 86%; Sp = 28% (cortical atrophy) (DAT vs normal controls) *Fazekas, et al., 1989*
PET vs MRI		x	PET: Se = 97%; Sp = 84% (qualitative) MRI: Se = 92%; Sp = 60% (ventricular atrophy) (DAT vs normal controls) *Fazekas, et al., 1989*
PET vs SPECT		x	PET: Se = 80%; Sp = 100% (typical functional pattern) SPECT: Se = 80%; Sp = 65% (typical functional pattern) (DAT vs normal controls and vascular dementia) *Mielke, et al., 1994*

Se = sensitivity; Sp = specificity
AD, Alzheimers diseae
MRI, magnetic resonance imaging
CT, computed tomography
SPECT, single photon emission computed tomography
DAT, dementia of the Alzheimers type

VIII. REFERENCES

A. General

Banta HD, Luce BR. A system for health care technology assessment. in: *Health Care Technology and its Assessment*. Oxford University Press, 1993, p. 61-82.

Chalmers, TC. PET scans and technology assessment. *Journal of the American Medical Association* 1988;260:2713-5.

Cooper LS, Chalmers TC, McCally M, Berrier J, Sacks HS.The poor quality of early evaluations of magnetic resonance imaging. *Journal of the American Medical Association* 1988;259:3277-80.

Egglin TKP, Feinstein AR. Context bias: a problem in diagnostic radiology. *JAMA* 1996; 276:1752-5.

Eysenck HJ. Meta-analysis and its problems. *British Medical Journal* 1994;309:789-92.

Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Medical Decision Making* 1991; 11:88-94.

Goodman C, Snider GL, Flynn K: *Primer: Technology Assessment In VA*. Management Decision and Research, Boston MA; VA Health Services Research and Development Service, Washington DC, 1996

Hasselblad V, Hedges LV. Meta-analysis of screening and diagnostic tests. *Psychological Bulletin* 1995;117(1):167-78.

Haynes RB, Sackett D, editors: Purpose and procedure (abbreviated). *Evidence-Based Medicine* 1995;1:2.

Hoffman RM, Kent DL, Deyo RA. Diagnostic accuracy and clinical utility of thermography for lumbar radiculopathy: a meta-analysis. *Spine* 1992;16:623-8.

Huston P, Moher D. Redundancy, disaggregation, and the integrity of medical research. *Lancet* 1996;347:1024-26.

Institute of Medicine. *Assessing Medical Technologies*. National Academy Press, Washington DC, 1985.

Kent DL, Haynor DR, Longstreth WT, Larson EB. The clinical efficacy of magnetic resonance imaging in neuroimaging. *Annals of Internal Medicine* 1994;120:856-71.

Kent DL, Larson EB. Disease, level of impact, and quality of research methods: three dimensions of clinical efficacy assessment applied to magnetic resonance imaging. *Investigative Radiology* 1992;27:245-54.

Kippenhan JS, Barker WW, Pascal S, Nagel J, Duara R. Evaluation of a neural-network classifier for PET scans of normal and Alzheimer's disease subjects. *Journal of Nuclear Medicine 1992*; 33:1459-67.

Kumar A, Schapiro MB, Grady C, Haxby JV, Wagner E, Salerno JA, et al. High-resolution PET studies in Alzheimer's disease. *Neuropsychopharmacology* 1991;4:35-46.

Light RJ, Pillemer DB. *Summing Up: The Science of Reviewing Research.* Harvard University Press, Cambridge, Massachusetts 1984.

Links JM, Devous MD. Detection and comparison of patterns in images. *Journal of Nuclear Medicine* 1994;35:16-17.

McMaster University Health Sciences Centre Department of Clinical Epidemiology and Biostatistics. How to read clinical journals: II. To learn about a diagnostic test. Canadian Medical Journal 1981;124:703-10.

Mulrow CD. The medical review article: state of the science. *Annals of Internal Medicine* 1987; 106:485-8.

Mulrow CD. Rationale for systematic reviews.*British Medical Journal* 1994;309:597-9.

Office of Technology Assessment: *Strategies for Medical Technology Assessment.* Pub. No. OTA-H-181. U.S. Government Printing Office, Washington DC, 1982.

Powers WJ, Berg L, Perlmutter JS, Raichle ME. Technology assessment revisited: Does positron emission tomography have proven clinical efficacy? *Neurology* 1991;41:1339-40.

Slavin RE. Best-evidence synthesis: an alternative to meta-analytic and traditional reviews. *Educational Researcher* 1986;15:5-11.

Slavin RE. Best evidence synthesis: an intelligent alternative to meta-analysis. *Journal of Clinical Epidemiology* 1995;1:9-18.

Thornbury JR, Kido DK, Mushlin AI, Phelps CE, Mooney C, Fryback DG. Increasing the scientific quality of clinical efficacy studies of magnetic resonance imaging. *Investigative Radiology* 1991;26:829-35.

B. Head and Neck Cancer Studies (Table 1)

Benchaou M, Lehmann W, Slosman DO, Becker M, Lemoine R, et al. The role of FDG-PET in the preoperative assessment of N-staging in head and neck cancer. *Acta Otolaryngologica* 1996;116:332-5.

Braams JW, Pruim J, Freling NJM, Nikkels PGJ, Roodenburg JLN, Boering G, et al. Detection of lymph node metastases of squamous cell cancer of the head and neck with FDG-PET and MRI. *Journal of Nuclear Medicine* 1995;36:211-6.

Lapela M, Grenman R, Kurki T, Joensuu,H, Leskinen ,Lindholm P, et al. Head and neck cancer: detection of recurrence with PET and 2-[F-18]Fluoro-2-deoxy-D-glucose. *Radiology* 1995; 197:205-11.

Laubenbacher C, Saumweber D, Wagner-Manslau C, Kau RJ, Herz M, Avril N, et al. Comparison of fluorine-18-fluorodeoxyglucose PET, MRI and endoscopy for staging head and neck squamous-cell carcinomas. Journal of Nuclear Medicine 1995;36:1747-57.

McGuirt WF, Williams DW, Keyes JW, Greven KM, Watson NE, Geisinger KR, Cappellari, JO. A comparative diagnostic study of head and heck nodal metastases using positron emission tomography. *Laryngoscope* 1995;105:373-7.

Rege S, Maass A, Chaiken L, Hoh CK, Choi Y, Lufkin R, et al. Use of positron emission tomography with fluorodeoxyglucose in patients with extracranial head and neck cancers. *Cancer* 1994;73:3047-58.

C. Colorectal Cancer Studies (Table 2)

Corman ML, Galandiuk S, Block GE, Prager ED, Weiner GJ, Kahn D, et al. Immunoscintigraphy with ^{111}In-satumomab pendetide in patients with colorectal adenocarcinoma: performance and impact on clinical management. *Diseases of the Colon and Rectum* 1994;37:129-37.

Falk PM, Gupta NC, Thorson AG, Frick MP, Bowman BM, Christensen MA, et al. Positron emission tomography for preoperative staging of colorectal carcinoma. *Diseases of the Colon and Rectum* 1994;37:153-6.

Ito K, Kato T, Tadokoro M, Ishiguchi T, Oshima M, Ishigaki T, et al. Recurrent rectal cancer and scar: differentiation with PET and MR imaging. *Radiology* 1992;182:549-52.

Hawes RH. New staging techniques: endoscopic ultrasound. *Cancer* 1993;71:4207-13.

Hernandez-Socorro CR, Guerra C, Hernandez-Romero J, Rey A, Lopez-Facal P, Alvarez-Santullano V. Colorectal carcinomas: diagnosis and preoperative staging by hydrocolonic sonography. *Surgery* 1995;117:609-15

Lai DTM, Fulham M, Stephen MS, Chu K-M, Solomon M, Thompson JF, et al. The role of whole-body positron emission tomography with [^{18}F]fluorodeoxyglucose in identifying operable colorectal cancer metastases to the liver. *Archives of Surgery* 1996;131:703-7.

Nattinger AB. Colon Cancer Screening and Detection. In Panzer RJ, Black ER, Griner PF, eds. *Diagnostic Strategies for Common Medical Problems*. American College of Physicians, Philadelphia, 1991.

Panzer RJ. Hepatic Metastases. In Panzer RJ, Black ER, Griner PF, eds. *Diagnostic Strategies for Common Medical Problems*. American College of Physicians, Philadelphia, 1991.

Rafaelsen SR, Kronborg O, Larsen C, Fenger C. Intraoperative ultrasonography in detection of hepatic metastases from colorectal cancer. *Diseases of the Colon and Rectum* 1995;38:355-60.

Schiepers C, Penninckx F, De Vadder N, Mercks E, Mortelmans L, Bormans G, et al. Contribution of PET in the diagnosis of recurrent colorectal cancer: comparison with conventional imaging. European Journal of Surgical Oncology 1995;21:517-22.

Schlag P, Lehner B, Strauss LG, Georgi P, Herfarth C. Scar or recurrent rectal cancer: positron emission tomography is more helpful than immunoscintigraphy. *Archives of Surgery* 1989; 124:197-200.

Stark DD, Wittenberg J, Butch RJ, Ferrucci JT. Hepatic metastases: randomized, controlled comparison of detection with MR imaging and CT. Radiology 1987;165:399-406.

Strauss LG, Clorius JH, Schlag P, Lehner B, Kimmig B, Egenhart R, et al. Recurrence of colorectal tumors: PET evaluation. *Radiology* 1989;170:329-32.

Vitola JV, Delbeke D, Sandler MP, Campbell MG, Powers TA, Wright JK, et al. Positron emission tomography to stage suspected metastatic colorectal carcinoma to the liver. *American Journal of Surgery* 1996;171:21-6.

Yonakura Y, Benua RS, Brill AB, Som P, Yeh SDJ, Kemeny NE, et al. Increased accumulation of 2-deoxy-2-[18F]fluoro-D-glucose in liver metastases from colon cancer. Journal of Nuclear Medicine 1982;23:1133-7.

D. Breast Cancer Studies (Table 3)

Adler LP, Crowe JP, Al-Kaisi NK,Sunshine JL. Evaluation of breast masses and axillary lymph nodes with [f-18] 2-deoxy-2-fluoro-D-glucose PET. *Radiology*. 1993;187:743-50.

Avril N, Dose J, Jänicke F, Ziegler S, Römer W, Weber W, et al. Assessment of axillary lymph node involvement in breast cancer patients with positron emission tomography using radiolabeled 2-(fluorine-18)-fluoro-2-deoxy-D-glucose. *Journal of the National Cancer Institute.* 1996;88(17):1204-9.

Avril N, Dose J, Jänicke F, Bense S, Ziegler S, Laubenbacher C, et al. Metabolic characterization of breast tumors with positron emission tomography using F-18 fluorodeoxyglucose. *Journal of Clinical Oncology*. 1996;14:1848-57.

Nieweg OE, Kim EE, Wong WH, Broussard WF, Singletary SE, Hortobagyi GN, et al. Positron emission tomography with fluorine-18-deoxyglucose in the detection and staging of breast cancer. *Cancer.* 1993;71:3920-5.

Scheidhauer K, Scharl A, Peitrzyk U, Wagner R, Göhring UJ, Schomäcker K, et al. Qualitative [18F]FDG positron emission tomography in primary breast cancer: clinical relevance and practicability. *European Journal of Nuclear Medicine*. 1996;23(6):618-23.

E. Lung Cancer Studies (Table 4)

Chin R, Ward R, Keyes JW, Choplin RH, Reed JC, Wallenhaupt, *et al.* Mediastinal staging of non-small-cell lung cancer with positron emission tomography. *American Journal of Respiratory and Critical Care Medicine* 1995; 152:2090-6.

Inoue T, Kim EE, Komaki R, Wong FCL, Bassa P, Wong W, *et al.* Detecting recurrent or residual lung cancer with FDG-PET. *Journal of Nuclear Medicine* 1995; 36:788-93.

Knight SB, Delbeke D, Stewart JR, Sandler MP. Evaluation of pulmonary lesions with FDG-PET: Comparison of findings in patients with and without a history of prior malignancy. *Chest* 1996;109:982-8.

Kubota K, Matsuzawa T, Fujiwara T, Ito M, Hatazawa J, Ishiwata K, *et al.* Differential diagnosis of lung tumor with positron emission tomography: a prospective study. *Journal of Nuclear Medicine* 1990; 31:1927-33.

Patz EF, Lowe VJ, Hoffman JM, Paine SS, Harris LK, Goodman PC. Persistent or recurrent bronchogenic carcinoma: detection with PET and 2-[F-18]-2-deoxy-D-glucose. *Radiology* 1994; 191:379-82.

Patz EF, Lowe VJ, Goodman PC, Herndon J. Thoracic nodal staging with pet imaging with ^{18}FDG in patients with bronchogenic carcinoma. *Chest* 1995; 108:1617-21.

Sazon DAD, Santiago SM, Soo Hoo GW, Khonsary A, Brown C, Mandelkern M, *et al*. Fluorodeoxyglucose-positron emission tomography in the detection and staging of lung cancer. *American Journal of Respiratory and Critical Care Medicine* 1996; 153:417-21.

Scott W, Schwabe JL, Gupta NC, Dewan NA, Reeb SD, Sugimoto JT, *et al*. Positron emission tomography of lung tumors and mediastinal lymph nodes using F-18-fluorodeoxyglucose. *Annals of Thoracic Surgery* 1994; 58:698-703.

Scott WJ, Gobar LS, Terry JD, Dewan NA, Sunderland JJ. Mediastinal lymph node staging of non-small-cell lung cancer: a prospective comparison of computed tomography and positron emission tomography. *The Journal of Thoracic and Cardiovascular Surgery* 1996; 111:642-8.

Slosman DO, Spiliopoulos A, Couson F, Nicod L, Louis O, Lemoine R, *et al*. Satellite PET and lung cancer: a prospective study in surgical patients. *Nuclear Medicine Communications* 1993; 14:955-61.

Valk PE, Pounds TR, Hopkins DM, Haseman MK, Hofer GA, Greiss HB, *et al.* Staging non-small cell lung cancer by whole-body positron emission tomographic imaging. *Annals of Thoracic Surgery* 1995; 60:1573-82.

Wahl R , Quint LE, Greenough RL, Meyer CR, White RI, Orringer MB. Staging of mediastinal non-small cell lung cancer with FDG PET, CT, and fusion images: preliminary prospective evaluation. *Radiology* 1994; 191:371-7.

F. Solitary Pulmonary Nodules (Table 5)

Bury T, Dowlati A, Paulus P, Corhay JL, Benoit T, Kayembe JM, *et al.* Evaluation of the solitary pulmonary nodule by positron emission tomography imaging. *European Respiratory Journal.* 1996;9:410-14.

Dewan NA, Reeb SD, Gupta NC, Gobar LS, Scott WJ. PET-FDG imaging and transthoracic needle lung aspiration biopsy in evaluation of pulmonary lesions: a comparative risk-benefit analysis. *Chest.* 1995;108:441-6.

Duhaylongsod FG, Lowe VJ, Patz EF, Vaughn AL, Coleman RE, Wolfe WG. Detection of primary and recurrent lung cancer by means of f-18 fluorodeoxyglucose positron emission tomography (FDG PET). *The Journal of Thoracic and Cardiovascular Surgery*. 1995;110:130-140. (b)

Gupta NC, Maloof J, Gunel E. Probability of malignancy in solitary pulmonary nodules using fluorine-18-FDG and PET. *The Journal of Nuclear Medicine*. 1996;37:943-8.

G. Alzheimer s Disease (Table 6)

Burdette JH, Minoshima S, Vander Borght T, Tran DD, Kuhl DE. Alzheimer disease: improved visual interpretation of PET images by using three-dimensional stereotaxic surface projections. Radiology 1996;198:837-43.

Fazekas F, Alavi A, Chawluk JB, Zimmerman RA, Hackney D, Bilaniuk L, et al. Comparison of CT, MR, and PET in Alzheimer's dementia and normal aging. *Journal of Nuclear Medicine* 1989; 1607-15.

Herholz K, Perani D, Salmon E, Granck G, Fazio F, Heiss WD, et al. Comparability of FDG PET studies in probable Alzheimer's disease. *Journal of Nuclear Medicine* 1993;34:1460-6.

Jobst KA, Hindley NJ, King E, Smith AD. The diagnosis of Alzheimer's disease: a question of image? *Journal of Clinical Psychiatry* 1994;55(11, suppl):22-31.

Kippenhan JS, Barker WW, Nagel J, Grady C, Duara, R. Neural-network classification of normal and Alzheimer's disease subjects using high-resolution and low-resolution PET cameras. *Journal of Nuclear Medicine* 1994;35:7-15.

Mielke R, Pietrzyk U, Jacobs A, Fink, GR, Ichimiya A, Kessler J, et al. HMPAO SPET and FDG PET in Alzheimer's disease and vascular dementia: comparison of perfusion and metabolic pattern. *European Journal of Nuclear Medicine* 1994;1053-60.

Salmon E, Sadzot B, Mazuet P, Degueldre C, Lemaire C, Rigo P, et al. Differential diagnosis of Alzheimer's disease with PET. *Journal of Nuclear Medicine* 1994;35:391-98.

Van Gool WA, Walstra GJM, Teunisse S, Van der Zant FM, Weinstein HC, Van Royen EA. Diagnosing Alzheimer's disease in elderly, mildly demented patients: the impact of routine single photon emission computed tomography. *Journal of Neurology* 1995;242:401-5.

Appendix 1

Advisory Committee to the PET Assessment

Appendix 1

Advisory Committee to the PET Assessment

Marguerite Hays, M.D. *(Chair)*
ACOS Research Service
Palo Alto VAMC
3801 Miranda Avenue
Palo Alto, CA 94304

Martin Charns, D.B.A.
Director, MDRC (152M)
150 South Huntington Ave.
Boston, MA 02130

Daniel Deykin, M.D.
Director, Health Services Research & Development Service, VHA (*Retired)*

James Fletcher, M.D.
Chief, Nuclear Medicine Service (115)
VAMC
915 North Grand Blvd.
St. Louis, MO 63106

Clifford Goodman, Ph.D.
4501 Connecticut Ave., N.W. #823
Washington, D.C. 20008

Milton Gross, M.D.
Program Dir., Nuclear Medicine Service (115)
VAMC
2215 Fuller Road
Ann Arbor, MI 48105

Thomas Holohan, M.D.
Chief, Office of Patient Care Services (11)
810 Vermont Avenue, N.W.
Washington, D.C. 20420

Steven Hotta, M.D.
Project Manager, PET
Center for Health Care Technology
Suite 309
6000 Executive Blvd.
Rockville, MD 20852

Steven Larson, M.D.
Chief, Nuclear Medicine Service (S212C)
Memorial Sloane Kettering Hospital
1275 York Avenue
New York, NY 10021

H. William Strauss, M.D.
Stanford University School of Medicine
Division of Nuclear Medicine
Room H0101
Route 8
Stanford, CA 94305-5281

Appendix 2

Assessing Diagnostic Technologies

Author: Karen Flynn, D.D.S., M.S., *Manager, MDRC Technology Assessment Program*

Appendix 2

Assessing Diagnostic Technologies[1]

This report summarizes the approach of the Management Decision and Research Center Technology Assessment Program to evaluating diagnostic technologies. The Program relies on this approach for both its major assessments (such as that of positron emission tomography and picture archiving and communications systems) and the reports issued in response to requests of the Technology Assessment Information Service.

The report is intended to supply readers with an understanding of the basic analytic tools that would be used in evidence-based decisions to perform a diagnostic test and to interpret its results. A similar analytic process can be applied to policy decisions regarding acquiring a diagnostic technology for a hospital and offering it for use in the diagnostic strategy for specific diseases.

1 This appendix was published as a separate report entitled "*MTA94-001-01: Assessing Diagnostic Technologies, July, 1996.*"

I. BACKGROUND

Sackett, et al. (1991) define ***diagnosis*** as "..the crucial process that labels patients and classifiestheir illnesses, that identifies (and sometimes seals!) their likely fates or prognoses, and that propels us toward specific treatments in the confidence (often unfounded) that they will do more good than harm."

The rationale for rigorously assessing diagnostic tests has been discussed at length (Sox, et al., 1989), and a number of imaging tests have been subjected to a high degree of scrutiny [e.g. magnetic resonance imaging for multiple sclerosis (Mushlin, et al., 1993; Phelps and Hutson, 1995) and thermography for lumbar radiculopathy (Hoffman, et al., 1992)]. Rigorous reviews of evidence on MRI (Kent and Larson, 1988; Kent and Larson, 1992; Kent, et al., 1994) concluded that many accuracy and utility questions remained unanswered due to lack of methodologic rigor. The lack of rigor in the early clinical studies of MRI further confirms the need to study diagnostic technologies carefully as they first move into clinical use (Cooper, et al., 1988; Sheps, 1988; Beam, et al., 1991).

This scrutiny reflects the recognition that health care resources are finite. It also reflects the recent movement from intuitive or informal clinical decision making based on the experience of individual practitioners to a more formal process in which evidence from studies of groups of similar patients supplements practitioners' judgment. The latter model of decision making, now referred to as "evidence-based medicine", requires critical appraisal of the literature and quantitative decision support. The orientation of diagnostic technology assessment around principles of evidence-based medicine reflects the mission of the MDRC Technology Assessment Program to promote evidence-based decision making within VA, through its own efforts and through those of its affiliated San Antonio Cochrane Center.

Diagnostic tests are performed in clinical practice when the information available from the history, physical examination, and any previous testing is considered insufficient to address the questions at hand (Black, et al., 1991). The decision to perform a test is made on the assumption that the results will appreciably reduce the uncertainty surrounding a given question and significantly change the pretest probability of disease. In other words, the overriding criterion for when to use a diagnostic test should be the usefulness of a given piece of diagnostic information to the clinician and to the patient. A useful diagnostic test does several things: it provides an accurate diagnosis, supports the application of a specific efficacious treatment, and ultimately leads to a better clinical outcome for the patient (Sackett, et al., 1991).

Studies to determine the safety, efficacy, and outcomes of diagnostic tests require careful attention to principles of design and potential sources of bias if they are to provide valid and useful information to clinicians, patients, and policy makers; judging the quality of such studies and their applicability to clinical decision making in specific situations is central to evidence-based practice. The following brief overview of some of the issues involved in assessing diagnostic tests will outline the design of studies used to evaluate the accuracy of diagnostic tests and introduce measurements of diagnostic test accuracy.

II. CONDUCTING STUDIES TO EVALUATE DIAGNOSTIC TEST ACCURACY

Studies that measure the accuracy of diagnostic tests are difficult to perform. Several authors (Riegelman and Hirsch,1989; Sackett, et al., 1991) have defined the kinds of studies that provide valid estimates of the accuracy of diagnostic tests. Others (Begg, 1987) have outlined potential sources of bias in such studies. More recently, Egglin and Feinstein (1996) have documented that sensitivity and specificity of subjectively interpreted tests (which would include qualitatively interpreted PET images in cancer patients) are biased by their context and by the prevalence of disease in recently observed cases; “context bias” is likely to skew quantitative mesaures of test performance from case series with high disease prevalence.

The following guides to designing and reporting a diagnostic test evaluation were adapted from those proposed by Sackett, et al., and Riegelman and Hirsch. They define a study that avoids potentially biased measurements of test accuracy, and that provides guidance on the usefulness of the test:

1) An independent, “blind” comparison with a “gold standard” of diagnosis is used.

2) The diagnostic test is evaluated in a patient sample that includes an appropriate spectrum of mild and severe, treated and untreated disease, plus individuals with different but commonly confused disorders.

3) A representative group of individuals without the disease is included. Ideally, as many diseased individuals as disease-free individuals defined by the gold standard are chosen. Although tests of statistical significance are rarely applied to assessments of diagnostic tests, this 50-50 split in the study population would give the greatest statistical power for a given sample size.

   Thornbury, et al. (1991), review the literature on sample sizes in evaluating the diagnostic accuracy of MRI, and provide guidelines for similar studies of other technologies (although absolute sample size requirements depend on the prevalence or pre-test probability of disease in the study sample and the change in diagnostic accuracy that is hypothesized or estimated to be associated with the test under study):

   - to compare the diagnostic accuracy of MRI to a gold standard, 30 to 70 subjects (30 would provide rough estimates of sensitivity and specificity; 70 would allow estimation within 7% to 10%);

   - to compare the diagnostic impact of MRI versus traditional imaging, 10 to 150 patients would be needed (major differences can be detected with 10 to 20 cases; more subtle differences would require up to 150);

   - to compare the therapeutic impact and/or patient outcome impact of two imaging techniques, 20 to 500 cases (again, depending on the magnitude of the difference expected) would be required.

4) The setting for the evaluation, as well as the filter through which study patients passed, is adequately described.

5) The reproducibility of the test result (precision) and its interpretation (observer variation) are determined.

6) The term “normal” is sensibly defined as it applies to the test.

7) If the test is advocated as part of a cluster or sequence of tests, its individual contribution to the overall validity of the cluster or sequence is determined.

8) The tactics for carrying out the test are described in sufficient detail to permit their exact replication.

9) The utility of the test is determined. Criteria for interpreting reports of utility include:

- the appropriate role of the test is studied (e.g., as a replacement for, or an addition to, an existing test);
- all clinically relevant outcomes (e.g., delays in therapy, complications from an invasive test, psychologic impacts of test) are reported;
- appropriate patients (e.g., those with neither a very low nor a very high probability of having the disease) are tested;
- statistically significant results are also clinically important;
- the test is feasible in the setting in which it will be applied by the clinician interpreting the report;
- all patients who entered the study are accounted for at its conclusion.

Once a study has been appropriately designed to measure the accuracy of a diagnostic test, the results of the study are analyzed and presented in the literature. Several measures of accuracy are available.

III. MEASURES OF THE ACCURACY OF DIAGNOSTIC TESTS

Each diagnostic test has a set of characteristics that reflect the results expected in patients with and without disease. Most diagnostic tests are imperfect to some extent. There is usually an overlap of test results among patients with and without a specific disease, causing healthy individuals to occasionally be classified wrongly as diseased, and some diseased individuals to fail to be detected. Study data documenting the extent to which a test result accurately reflects reality can be analyzed in several ways.

An approach to analyzing test results is selected according to the number of categories into which the results may be placed (two categories, more than two categories. or continuous values) and the uses to which the diagnostic information will be put (ascertaining the presence of disease versus the severity of disease).

A. Is disease present or absent?

In its simplest form, the assessment of a diagnostic technology involves two dichotomies: disease present or absent (determined by applying the gold standard test), and diagnostic test result positive or negative (i.e., the test yields only two values or it yields a series of values and one is assigned to be the threshold between presence and absence of disease). When the presence or absence of a disease is at issue, sensitivity and specificity are the measures of accuracy used. These are frequently calculated using a 2 x 2 table:

Matrix for calculating the characteristics of a diagnostic test

		DISEASE	
		Present	**Absent**
TEST	**Positive**	**a** (true positive)	**b** (false positive)
	Negative	**c** (false negative)	**d** (true negative)

Accuracy = (a + d)/(a + b + c + d)

- *The proportion of all test results (both positive and negative) which are correct.*

Sensitivity = a/(a+c)

- *The proportion of people with the disease who have a positive test.*
- *A sensitive test will rarely miss people with the disease, and is usually chosen when there is an important penalty for missing the disease (i.e. when a dangerous but treatable condition is suspected), during the early stages of a workup when many possibilities are under consideration, or when the probability of a particular disease is low.*
- *A sensitive test is most helpful to the clinician when the results are negative; a negative result in a highly sensitive test rules out a disorder.*

Specificity = d/(b+d)

- *The proportion of people without the disease who have a negative test.*
- *A specific test will rarely misclassify people without the disease as diseased, and is used to confirm a diagnosis that has been suggested by other data.*
- *Highly specific tests are particularly needed when false positive results can harm the patient physically, emotionally, or financially.*
- *A specific test is most helpful when the result is positive; a positive result in a highly specific test rules in a disorder.*

For most tests, there is some overlap in findings for those within and without disease. Different cutoff points to define the presence of disease yield different pairs of sensitivity and specificity values. As the cutoff point used to define an abnormal result is made less extreme, sensitivity will improve and specificity will worsen.

Sensitivity and specificity are often considered to be independent of disease prevalence. However, they do change with changes in prevalence if the mix (mild versus severe disease) of patients with the target disorder varies with prevalence. For example, sensitivity would decrease if a diagnostic test that had been evaluated in a tertiary care center were applied in a community hospital where the target condition was both less common and less severe. Specificity would also decrease if, in the community hospital, more patients without the target condition received treatments that could induce false-positive results (Sackett, et al., 1991).

B. What is the severity of disease?

When diagnostic information falls into several categories or behaves as a continuous variable, comparison with the gold standard becomes more complex. Test results that are grouped into more than two categories can indicate severity of disease (in addition to its presence or absence, which now can be assigned to any of a series of cutoff points). The multiple cutoffs that separate disease from no disease create a corresponding number of true and false positive rates. Graphs of the relationship between the pairs of sensitivity and false positive rates (1 - sensitivity) are called receiver operating characteristic (ROC) curves. The likelihood that a test result is a true positive varies with the point on the ROC curve.

An ROC curve can be used to determine an optimal cutoff point, according to the purpose of the test (e.g. rule in disease or rule out disease). The optimal cut point when the pretest probability of disease is approximately 50% is that nearest the upper left corner of the curve. ROC curves can also be used to compare the usefulness of two different diagnostic tests for the same disorder: the one that encloses the larger area is more accurate. ROC analysis does not require *a priori* selection of a single decision threshold to use with a new test, and facilitates *a posteriori* selection of the optimum threshold prior to use of the test in routine clinical practice.

The kappa statistic, a measure of the degree of agreement that occurred between the diagnostic test and the gold standard over and above that which would have occurred by chance alone, can be used as a measure of test accuracy when there are more than two categories of test results. A kappa of 1.0 indicates perfect agreement; 0 indicates complete disagreement. A weighted kappa can take into account the degree of disagreement, generating a higher score when disagreements are close than when they are far apart.

Correlation coefficients describe the relationship between the continuous variable results of the test and the gold standard. When the diagnostic test result goes up, the gold standard also goes up (and the reverse). Correlation coefficients (r) approach 1.0 when the relationship is strong, and .0 when it is weak. Squaring the correlation coefficient yields a measure of the degree to which variation in gold standard results is explained by test results; r^2 values greater than 50% are generally considered respectable.

Analysis of variance, analysis of covariance, and multiple regression may occasionally be used to analyze diagnostic test accuracy. The choice of method depends on the number and continuous or categorical nature of the test results.

IV. INTERPRETING RESULTS AFTER AN ACCURATE TEST HAS BEEN SELECTED AND PERFORMED

Measures of diagnostic test accuracy are taken into account when a decision is made to order a test. Sensitivity and specificity are the most widely understood and facilitate choosing a test to rule in or rule out a diagnostic hypothesis (Sackett, et al., 1991).

However, a test’s accuracy is only one determinant of its clinical usefulness. Once the results of the test are available, the probability that the patient has the disease, given the results of the test (i.e. the posttest probability of disease), becomes more important. The largest gains from pre- to posttest probability occur when the pretest probability of the target disorder is 40 to 60% (Sackett, et al., 1991), or when the posttest probability crosses a threshold for deciding to initiate treatment (Sackett, et al., 1991; Black, et al., 1991).

Ways of revising the probability of disease based on test results (i.e. calculating the posttest probability of disease, or interpreting the test results) include Bayes' theorem, which extrapolates information from the 2 x 2 table in Figure 1:

Positive predictive value = a/(a+b)

The probability of disease in a person with an abnormal /positive test result.

Negative predictive value = d/(c+d)

The probability of not having the disease when the test result is negative.

Positive and negative predictive values vary with the pretest probability (or prevalence) of disease; as prevalence falls, positive predictive value falls along with it and negative predictive value rises. Sox, et al. (1989), note that the efficacy of a test is context dependent; it is not possible to properly interpret the meaning of a test result without taking into account what was known about the patient before the test.

Likelihood ratios are increasingly used to calculate posttest probability of disease (nomograms are available to simplify the conversion process), and are independent of pretest probability in most circumstances (Sackett, et al., 1991). The likelihood ratio describes the relative odds of an outcome, given a particular test result. Tests with dichotomous results will have two likelihood ratios (positive and negative) that reflect the relative odds of a condition being present after a positive or negative test.

Likelihood ratios can also be determined for each of several intervals across a full range of possible test results (multiple, rather than dichotomous, results). Finally, likelihood ratios can be used in sequential testing where the posttest odds from the first test become the pretest odds for the next test (Suchman and Dolan, 1991). Likelihood ratios can be calculated from the 2 x 2 table:

Likelihood ratio (positive) = sensitivity/(1 - specificity)

Likelihood ratio (negative) = (1 - sensitivity)/specificity

V. ANALYTIC FRAMEWORK FOR MDRC TECHNOLOGY ASSESSMENT PROGRAM SYSTEMATIC REVIEWS OF DIAGNOSTIC TEST LITERATURE

MDRC assessments and reviews will use four specific analytic frameworks in reviewing the published literature; the broad outlines of the assessment approach have already been introduced. While there is some overlap among the frameworks, each brings a unique set of conceptual tools to an evaluation of the literature.

A. What is the quality of the individual studies that were intended to measure the technology's characteristics (accuracy) as a diagnostic test?

Criteria sets for assessing the quality of a diagnostic test evaluation have been reviewed (Mulrow, et al., 1989). An accessible and straightforward set of criteria has more recently been defined for use in evidence-based medicine (Haynes and Sackett, 1995). Evidence-based medicine applies the best available evidence in clinical and other health care decisions. Conversely, evidence that is of insufficient quality to use as a basis for clinical or policy decisions is screened out by evidence-based medicine criteria.

The evidence-based medicine criteria for diagnostic tests will be applied to the individual studies cited in MDRC technology assessment reports. If the criteria are not met, the study will generally be considered insufficiently rigorous to provide the basis for patient care decisions. However, such studies often provide useful information on the technical characteristics of a diagnostic test, or may provide information necessary to subsequent diagnostic accuracy studies.

Evidence-based medicine criteria for evaluating studies of diagnosis

• Clearly identified comparison groups, ≥ 1 of which is free of the target disorder.

• Either an objective diagnostic standard (e.g. a machine-produced laboratory result) or a contemporary clinical diagnostic standard (e.g. a venogram for deep venous thrombosis) with demonstrably reproducible criteria for any subjectively interpreted component (e.g., report of better-than-chance agreement among interpreters).

• interpretation of the test without knowledge of the diagnostic standard result (i.e., blinding of test interpreter to results with diagnostic standard).

• Interpretation of the diagnostic standard without knowledge of the test result (i.e., blinding of diagnostic standard interpreter to results of test being evaluated).

As will be highlighted again below, documentation of test accuracy does not translate into documentation that the test is clinically useful. Sensitivity and specificity, while not as dependent on pretest probability of disease as predictive values, can be biased by differences in the mix of patients in the accuracy study and the patients on whom the test will be used in clinical practice (Sackett, et al., 1991). A published study that does not supply valid information needed to calculate posttest probability of disease (i.e. predictive values or likelihood ratios) would not assist clinicians in interpreting its results, or taking action based on those results.

Evidence-based criteria provide a broad quality screen for clinicians who are contemplating using a test in their own patients. A somewhat more detailed set of quality criteria, that expand on those of evidence-based medicine, have been used by the American College of Physicians in evaluations of the literature on magnetic resonance imaging (Kent, et al., 1994; Kent and Larson, 1992; Kent and Larson, 1988). These criteria are tabulated on the next page.

Methodologic quality of diagnostic accuracy studies

Grade	*Criteria*
A	**Studies with broad generalizability to a variety of patients and no significant flaws in research methods** • ≥ 35 patients with disease and ≥ 35 patients without disease (since such numbers yield 95% CIs whose lower bound excludes 0.90 if Se = 1) • patients drawn from a clinically relevant sample (not filtered to include only severe disease) whose clinical symptoms completely described • diagnoses defined by an appropriate reference standard • PET studies technically of high quality and evaluated independently of the reference diagnosis
B	**Studies with a narrower spectrum of generalizability, and with only a few flaws that are well described (and impact on conclusions can be assessed)** • ≥ 35 cases with and without disease • more limited spectrum of patients, typically reflecting referral bias of university centers (more severe illness) • free of other methods flaws that promote interaction between test result and disease determination • prospective study still required
C	**Studies with several methods flaws** • small sample sizes • incomplete reporting • retrospective studies of diagnostic accuracy
D	**Studies with multiple flaws in methods** • no credible reference standard for diagnosis • test result and determination of final diagnosis not independent • source of patient cohort could not be determined or was obviously influenced by the test result (work up bias) • opinions without substantiating data

B. Where does an individual study fall in the hierarchy of diagnostic efficacy?

Accurate estimation of the characteristics of a diagnostic test is one of the early steps in the assessment of that test. However, a complete assessment requires further research.

Fryback and Thornbury (1991) note that the localized view of the goal of diagnostic radiology would be that it provide the best images and the most accurate diagnoses possible. A more global view recognizes diagnostic radiology as part of a larger system of medical care whose goal is to treat patients effectively and efficiently. Viewed in this larger context, even high-quality images may not contribute to improved care in some instances, and images of lesser quality may be of great value in others. The point of the systematic view is to examine the ultimate value or benefit that is derived from any particular diagnostic examination.

Fryback and Thornbury (1991; 1992) present the most recent manifestation of an evolving hierarchical model for assessing the efficacy of diagnostic imaging procedures. Their model, with a list of the types of measures which appear in the literature at each level in the hierarchy, is presented in Table 1. As noted above, this assessment has adopted evidence-based medicine criteria as a requirement for assignment of studies to the "diagnostic accuracy" level of the hierarchy.

The table goes from the micro, or local level, at which the concern is the physical imaging process itself, to the societal efficacy level. The model stipulates that for a procedure to be efficacious at a higher level in the hierarchy it must be efficacious at the lower levels, but the reverse is not true; this asymmetry is often lost in research reports at Levels 1 and 2. Using this model, it is possible to follow the development of a diagnostic technology, and to align current research efforts with a particular level of development.

The diagnostic efficacy hierarchy is conceptually useful, but has some limitations as a guide to assessing the quality of individual studies. Judging the validity and generalizability of studies that address levels of the hierarchy beyond diagnostic accuracy requires additional criteria, which are discussed in the next section.

C. How strong is the evidence supporting a causal link between the use of the technology and improved outcomes of care?

The third analytic framework for the review of the literature will "grade" the available evidence for the degree to which it supports a causal link between the use of the technology and improved outcomes (i.e., Levels 4, 5, and 6 of the diagnostic efficacy hierarchy discussed in the section above). "Grading" evidence that is gathered in a comprehensive literature search according to its methodological rigor is an increasingly standard approach to health care technology assessment (Goodman, 1995).

Cook, et al. (1992) synthesize current thinking on the relative strength associated with the various study designs; this thinking is summarized in Table 2.

VI. SYSTEMATIC REVIEW PROTOCOL

The systematic reviews of the diagnostic test literature produced by the MDRC use a review protocol to guide the assignment of quality and diagnostic efficacy levels to studies. A typical protocol follows a defined sequence of steps:

---

1) Conduct MEDLINE and other database searches; retrieve full text articles that meet screening criteria:

   - English language articles reporting primary data and published in a peer review journal (not abstracts)
   - studies≥ 12 human subjects (not animal studies) with the disease of interest (sample sized defined by PET Advisory Committee)
   - studies using the radiopharmaceutical 2-[^{18}F]fluoro-2-D-glucose (FDG)

2) Apply screening criteria to bibliographies of retrieved articles as above, and retrieve additional articles.

3) Review full text articles and assign to level of Fryback and Thornbury (1991) diagnostic efficacy hierarchy.

Systematic review protocol, cont'd.

4) Assign to **technical efficacy** level of Fryback and Thornbury diagnostic efficacy hierarchy:

- uncontrolled studies
- feasibility studies
- correlation studies of glucose metabolic rate changes with treatment

Studies whose stated purpose is to define diagnostic accuracy but which report results in a way that measures of diagnostic accuracy cannot be duplicated or interpreted, or in which some patients entered are not accounted for, will also be assigned to the technical efficacy level.

5) Assign to **diagnostic accuracy efficacy** level:

- stated purpose is to define diagnostic accuracy , and clinically useful measures (Se/Sp) provided or can be calculated
- meets full or modified (case series with internal controls; blinding if image analysis qualitative) evidence-based medicine criteria
- determines optimal cutpoint from ROC analysis or applies previously determined optimal cutpoint

Caveats will be attached to reports of sensitivity and specificity reported for case series with internal controls if prevalence of severe disease is high.

6) Assign to **diagnostic thinking efficacy** level if meets evidence-based medicine criteria for evaluations of diagnostic tests and:

- numbers of subjects without target disorder ≥ numbers of cases with disorder (i.e. pretest probability of disease ≈ 50%)
- information useful in interpreting test results (i.e. converting pre- test probability of disease to post-test probability using predictive values or likelihood ratios) is provided or can be calculated from information in article.

Evidence-based medicine criteria for studies of diagnostic tests

- Clearly identified comparison groups, ≥ 1 of which is free of the target disorder.
- Either an objective diagnostic standard (e.g. a machine-produced laboratory result) or a contemporary clinical diagnostic standard (e.g. a venogram for deep venous thrombosis) with demonstrably reproducible criteria for any subjectively interpreted component (e.g., report of better-than-chance agreement among interpreters).
- interpretation of the test without knowledge of the diagnostic standard result.
- Interpretation of the diagnostic standard without knowledge of the test result.

Systematic review protocol, cont'd.

7) To further refine judgment of methodologic quality, grade **diagnostic accuracy or thinking efficacy** studies

Methodologic quality of diagnostic accuracy and diagnostic thinking efficacy studies*

Grade	*Criteria*
A	**Studies with broad generalizability to a variety of patients and no significant flaws in research methods** • ≥ 35 patients with disease and ≥ 35 patients without disease (since such numbers yield 95% CIs whose lower bound excludes 0.90 if Se = 1) • patients drawn from a clinically relevant sample (not filtered to include only severe disease) whose clinical symptoms completely described • diagnoses defined by an appropriate reference standard • PET studies technically of high quality and evaluated independently of the reference diagnosis
B	**Studies with a narrower spectrum of generalizability, and with only a few flaws that are well described (and impact on conclusions can be assessed)** • ≥ 35 cases with and without disease • more limited spectrum of patients, typically reflecting referral bias of university centers (more severe illness) • free of other methods flaws that promote interaction between test result and disease determination • prospective study still required
C	**Studies with several methods flaws** • small sample sizes • incomplete reporting • retrospective studies of diagnostic accuracy
D	**Studies with multiple flaws in methods** • no credible reference standard for diagnosis • test result and determination of final diagnosis not independent • source of patient cohort could not be determined or was obviously influenced by the test result (work up bias) • opinions without substantiating data

8) Assign to **therapeutic efficacy level** if meets evidence-based criteria for evaluations of diagnostic tests and/or:

- authors discuss how test results did change, or could have changed, treatment for the patients enrolled in the study
- % of times subsequent procedure avoided due to test results, % of times prospectively stated therapeutic plans changed post-test documented.

9) Assign to **patient outcome efficacy** level if patient outcomes with PET are compared to those without PET in a case-control study, cohort study, or randomized controlled trial and/or:

- change in quality adjusted survival or cost/quality adjusted life year gained documented.

10) Assign to **societal efficacy** level if both costs (from a societal perspective) and consequences (efficacy, effectiveness, or utility) determined for both PET and an alternative.

11) Evaluate quality of studies at each efficacy level; conduct meta analyses if appropriate.

12) Articles are excluded from the review if they:

- are duplicated or superseded by subsequent study (at the same level of the hierarchy and with the same purpose) from the same institution
- contain insufficient information to judge comparability of case and control groups, details of imaging protocol, whether visual or quantitative analysis of PET data used, or type of PET quantitative data analysis used.

Table 1: A Hierarchical Model of Efficacy for Diagnostic Imaging

Level	*Typical Measures of Analysis*	*Comments*
1. Technical efficacy	• Resolution of line pairs • Modulation transfer function • Gray-scale range • Amount of mottle • Sharpness	• Physical parameters describing technical imaging quality
2. Diagnostic accuracy efficacy	• Yield of abnormal or normal diagnoses in a case series • Diagnostic accuracy (% of correct diagnoses in case series) • Positive or negative predictive value in a case series • Sensitivity and specificity in a defined clinical setting • Measures of ROC curve height (d') or area under the curve A_z	• Joint function of images and observer • Also a function of clinician who requests diagnostic procedure, since selection controls specificity of test in clinical practice and sensitivity to the extent that it varies with the spectrum of disease.
3. Diagnostic thinking efficacy	• Number (%) of cases in series in which image judged "helpful" to making diagnosis • Entropy change in differential diagnosis probability distribution • Difference in clinicians' subjectively estimated diagnosis probabilities pre- to posttest information • Empirical subjective log-likelihood ratio for test positive and negative in a case series	• Inducing change in clinicians' diagnostic thinking is necessary prerequisite to impact on patients • Clinicians may value results which reassure them, but which do not change treatment decisions • Empirical methods to measure change in pretest diagnostic probabilities assumed by clinicians are probably best for determining the absence of diagnostic thinking efficacy, rather than estimating the magnitude of change in diagnostic thinking due to imaging information • Imaging examination result may influence clinician's diagnostic thinking, but have no impact on patient treatment

Level	*Typical Measures of Analysis*	*Comments*
4. Therapeutic efficacy	• Number (%) of times images judged helpful in planning management of the patient in a case series • % of times medical procedure avoided due to image information • % of times therapy planned pretest changed after imaging information was obtained (retrospectively inferred from patient records) • % of times clinicians' prospectively stated therapeutic choices changed after test information	• In situations where RCTs of decision making with and without the imaging information cannot be performed ethically or because of the momentum for using a particular procedure, asking Level 4 questions may be only efficacy study possible • Integrating negative information about a test from Level 3 and 4 studies may help to direct clinical use away from imaging tests that are not useful, or have been supplanted by other tests
5. Patient outcome efficacy	• % of patients improved with test compared with no test • Morbidity (or procedures) avoided with test • Change in quality-adjusted life expectancy • Expected value of test information in QALYS • Cost per QALY saved with imaging information	• Definitive answer re efficacy with respect to patient outcome requires RCT (involving withholding test from some patients) • RCTs may be associated with formidable statistical, empirical, and ethical problems and are justified only in carefully selected situations • Weaker evidence may be derived from case control studies or case series • Independent contribution of imaging to patient outcome may be small, requiring very large sample sizes • Decision analytic approach can be alternative to RCT, but the analyses may suffer from the same biases as their secondary data sources • Decision analyses can highlight critical pieces of information and guide future studies
6. Societal efficacy	• Cost-benefit analysis from societal viewpoint • Cost-effectiveness analysis from societal viewpoint • Cost-utility analysis from societal viewpoint	• Economic evaluations of evolving technologies do not provide definitive answers, since values and judgments play a significant role in interpretation of results • Cost-utility analyses imply at least Level 5 efficacy data or models

Adapted from Fryback and Thornbury, 1991

Abbreviations:
RCT, randomized clinical trial
ROC, receiver operating characteristic
QALY, quality adjusted life year

Table 2: **Judging the quality of individual studies:**
Classifications of study designs and levels of evidence
(when high quality meta analyses/overviews are not available)

Level	*Description*
I	**Randomized trials with low false-positive (alpha) and low false-negative (beta) errors (high power)** *• positive trial with statistically significant treatment effect (low alpha error)* *• negative trial that was large enough to exclude the possibility of a clinically important benefit (low beta error/high power; i.e. had a narrow confidence interval around the treatment effect, the lower end of which was greater than the minimum clinically important benefit)* *• meta analysis can be used to generate a pooled estimate of treatment efficacy across all high quality, relevant studies and can reveal any inconsistencies in results*
II	**Randomized trials with high false-positive (alpha) and/or high false negative (beta) errors (low power)** *• trial with interesting positive trend that is not statistically significant (high alpha error)* *• negative trial but possibility of a clinically important benefit (high beta error/low power; i.e. very wide confidence intervals around the treatment effect)* *• small positive trials with wide confidence intervals around the treatment effect, making it difficult to judge the magnitude of the effect* *• when Level II studies are pooled (through quantitative meta analysis), the aggregate effects may provide Level I evidence*
III	**Nonrandomized concurrent cohort comparisons between contemporaneous patients who did and did not (through refusal, noncompliance, contraindication, local practice, oversight, etc.) receive treatment** *• results subject to biases* *• Level III data can be subjected to meta analysis, but the result would not shift these data to another Level, and is not usually recommended*
IV	**Nonrandomized historical cohort comparison between current patients who did receive treatment (as a result of local policy) and former patients (from the same institution or from the literature) who did not (since at another time or in another institution different treatment policies prevailed).** *• results subject to biases, including those that result from inappropriate comparisons over time and space*
V	**Case series without control subjects** *• may contain useful information about clinical course and prognosis but can only hint at efficacy*

Source: Cook, et al., 1992

VII.REFERENCES

Bailar JC, Mosteller F, eds. *Medical Uses of Statistics*. Second Edition. Boston: NEJM Books, 1992.

Beam CA, Sostman HD, Zheng JY. Status of clinical MR evaluations 1985-1988: baseline and design for further assessments. *Radiology* 1991; 180:265-70.

Begg CB. Biases in the assessment of diagnostic tests. *Statistics in Medicine* 1987; 6:411-23.

Black ER, Panzer RJ, Mayewski RJ, Griner PF. Characteristics of diagnostic tests and principles for their use in quantitative decision making. in: Panzer, et al., eds. *Diagnostic Strategies for Common Medical Problems.* Philadelphia: American College of Physicians, 1991.

Black WC, Dwyer AJ. Local versus global measures of accuracy: an important distinction for diagnostic imaging. *Medical Decision Making* 1990; 10:266.

Cook DJ, Guyatt G, Laupacis, A, Sackett DL. Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest* 1992; 102(4, Supplement):305S-11S.

Cooper LS, Chalmers TC, McCally M, Berrier J, Sacks HS. The poor quality of early evaluations of magnetic resonance imaging. *Journal of the American Medical Association* 1988; 259:3277-80.

Egglin TKP, Feinstein AR. Context bias: a problem in diagnostic radiology. *JAMA* 1996;276:1752-5.

Fletcher RH, Fletcher SW, Wagner EH: *Clinical Epidemiology*. Second edition. Baltimore: Williams and Wilkins, 1988.

Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Medical Decision Making* 1991; 11:88-94.

Guyatt GH, Sackett DL, Sinclair JC, Hayward R, Cook DJ, Cook RJ, for the Evidence-Based Medicine Working Group. Users' guides to the medical literature: IX. A method for grading health care recommendations. *Journal of the American Medical Association* 1995; 274:1800-04.

Haynes RB. Tracking down and reading the literature to learn about diagnostic tests. in: Panzer, et al., eds., *Diagnostic Strategies for Common Medical Problems.* Philadelphia: American College of Physicians, 1991.

Haynes RB, Sackett D, eds. Purpose and procedure (abbreviated). Evidence-Based Medicine 1995, 1:2.

Hoffman RM, Kent DL, Deyo RA. Diagnostic accuracy and clinical utility of thermography for lumbar radiculopathy: a meta-analysis. *Spine* 1992; 16:623-8.

Jaeschke R, Guyatt G, Sackett DL for the Evidence-Based Medicine Working Group. Users' guide to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid? *Journal of the American Medical Association* 1994; 271(5):389-91.

Jaeschke R, Guyatt G, Sackett DL for the Evidence-Based Medicine Working Group. Users' guide to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? *Journal of the American Medical Association* 1994; 271(9):703-7.

Kent DL, Haynor DR, Longstreth WT, Larson EB. The clinical efficacy of magnetic resonance imaging in neuroimaging. *Annals of Internal Medicine* 1994; 120:856-71.

Kent DL, Larson EB. Magnetic resonance imaging of the brain and spine: is clinical efficacy established after the first decade? *Annals of Internal Medicine* 1988; 108:402-24.

Kent DL, Larson EB. Disease, level of impact, and quality of research methods: three dimensions of clinical efficacy assessment applied to magnetic resonance imaging. *Investigative Radiology* 1992; 27:245-54.

Mulrow CD, Linn WD, Gaul MK, Pugh JA. Assessing quality of a diagnostic test evaluation. *Journal of General Internal Medicine* 1989; 4:288-95.

Mushlin AI, Detsky AS, Phelps CE, O'Connor P, Kido DK, Kucharczyk W, et al. for the Rochester-Toronto Magnetic Resonance Imaging Study Group. The accuracy of magnetic resonance imaging in patients with suspected multiple sclerosis. *Journal of the American Medical Association* 1993; 269:3146-51.

Panzer RJ, Black ER, Griner PF. Interpretation of diagnostic tests and strategies for their use in quantitative decision making. in: Panzer, et al., eds., *Diagnostic Strategies for Common Medical Problems.* Philadelphia: American College of Physicians, 1991.

Riegelman RK, Hirsch RP. *Studying a Study and Testing a Test: How to Read the Medical Literature.* Second Edition. Boston: Little, Brown and Company, 1989.

Sackett DL, Haynes RB, Guyatt GH, Tugwell P. *Clinical Epidemiology: A Basic Science for Clinical Medicine.* Second Edition. Boston: Little, Brown and Company, 1991.

Sox H, Stern S, Owens D, Abrams HL. Monograph of the Council on Health Care Technology, Institute of Medicine: *Assessment of Diagnostic Technology in Health Care*. Washington DC: National Academy Press, 1989.

Suchman AL, Dolan JG. Odds and likelihood ratios. in: Panzer, et al., eds. *Diagnostic Strategies for Common Medical Problems.* Philadelphia: American College of Physicians, 1991.

Thornbury JR, Kido DK, Mushlin AI, Phelps CE, Mooney C, Fryback DG. Increasing the scientific quality of clinical efficacy studies of magnetic resonance imaging. *Investigative Radiology* 1991; 26:829-35.

Woolf SH, Battista RN, Anderson GM, Logan AG, Wang E, the Canadian Task Force on the Periodic Health Examination. Assessing the clinical effectiveness of preventive maneuvers: analytic principles in reviewing evidence and developing clinical practice recommendations. *Journal of Clinical Epidemiology* 1990; 43:891-905.

SECTION VIII. GLOSSARY

**note: words in italics have been defined elsewhere in the glossary*

Accuracy: the proportion of all test results (both positive and negative) that are correct; results close to the true measure of the biologic phenomenon; accuracy depends on the *validity* and *precision* of the study.

Alpha: *false-positive* error; also *Type I* error.

Bayes' Theorem (Bayesian analysis): a mathematical model used to calculate *post-test probabilities* for diagnostic tests and procedures (i.e., *pre-test probability of disease X likelihood ratio for the diagnostic test result = post-test probability of disease*); also expressed in terms of the *odds* of disease before knowing the symptom and after knowing the symptom; commonly applied to clinical decision analysis to estimate the probability of a diagnosis given some symptom or test result (i.e., *post-test probability*).

Beta: *false-negative* error; see also *Type II* error.

Bias: a type of systematic error; any effect at a stage of investigation or inference tending to produce results that depart systematically from the true values.

Case: a person in the study group who has the disease or characteristic of interest.

Case-control study: a type of retrospective, nonexperimental study design especially useful for studies of rare diseases whereby first the cases and a similar referent sample without the disease (i.e., "controls") are identified by census, after which the researcher looks back in time to determine the frequency of exposure to the risk factor(s) of interest.

Case-mix: features of a study population that increase the risk of a bad outcome or influence the choice of treatment (e.g., severity of disease, coexisting conditions); such features must be taken into account when assessing treatment outcomes.

Case series: a type of *nonexperimental study* design in which an investigator reports a group or series of cases with the characteristic of interest; although among the most common, it represents the weakest of studies designed to establish causation.

Chance: something that happens unpredictably without intervention.

Clinical pathway: a multidisciplinary set of guidelines and outcome targets for managing the overall care of a particular patient or patient group; often used as a method of quality assurance and a cost-reduction strategy for patients in particular diagnostic-related groups.

Clinical practice guidelines: a systematically developed set of parameters for one or more specific clinical circumstances used to assist practitioners in health care decision making.

Clinical significance: the effect that a technology or intervention has which is meaningful to patients and/or health care providers; however, it may or may not have *statistical significance.*

Cochrane Center (in San Antonio): part of the *Cochrane Collaboration* funded by VA *HSR&D* Service through the *MDRC Technology Assessment Program* to provide an information clearinghouse for all U.S. *Cochrane Collaboration* participants and anyone interested in obtaining information about the Collaboration; its focus includes development and coordination of training programs for those preparing and maintaining systematic reviews, and building a database of trials in hypertension treatment.

Cochrance Collaboration: an international, non-profit endeavor that aims to prepare, maintain, and disseminate systematic reviews of health care comprising Centers, Reviews Groups, Fields, Method Groups, and a Consumer Network; the Centers support and facilitate the work of the Collaboration.

Cohort study: follow-up or longitudinal study; a prospective, *nonexperimental study* in which a defined subset of the population is followed for a defined period to compare the outcomes in a group of patients that received an exposure or intervention to a similar group that did not receive the exposure or intervention; a weaker study design than a *randomized clinical trial* used to establish a casual link between the intervention and outcome of interest, but may be the most feasible approach to answer the questions of interest.

Confounding: distortion of the effect of an exposure on risk resulting from a *confounding variable.*

Confounding variable: a factor that is unequally distributed among the exposed and unexposed and independently affects the risk of developing the disease; "confounder."

Control group: referent group; a group of study subjects to which the effects of an intervention given to the treatment group is compared and who, with the exception of the intervention, resemble the treatment group as closely as possible.

Continuous variable: quantitative data that may take on fractional values (e.g., height, weight, serum cholesterol).

Correlation coefficient: a numeric measure between -1 and 1 expressing the observed linear association between two variables; expressed as r, the value r=0 indicates a nonlinear relationship between the two variables.

Cost-benefit analysis: an economic analysis which expresses the outcome of interest (or the benefit) in terms of currency (e.g., loss in net earnings due to death or disability).

Cost-effectiveness analysis: an economic analysis which compares the outcome of decision options in terms of their monetary cost per unit of health *outcome* achieved; health *outcomes* are measured in terms of health status.

Cost-utility analysis: an economic analysis which incorporates relative social value or preferences into the health *outcome* considered: often expressed as a monetary cost per "quality-adjusted life year."

Decision analysis: a systematic quantitative approach used to assess the relative value of one or more clinical approaches; often expressed graphically in the form of a decision tree.

Diagnosis: the process of determining one's health status and the factors responsible for producing it.

Diagnostic accuracy: a characteristic of *diagnostic test efficacy* describing the proportion of all test results that are correct.

Diagnostic impact: a characteristic of *diagnostic test efficacy* describing the effect of test results on diagnosis (i.e., the change from *pretest probability* to *posttest probability*); may not necessarily affect treatment decisions.

Diagnostic test efficacy: the impact and usefulness of a diagnostic test expressed in terms of its technical properties, *diagnostic accuracy*, or its impact on *diagnosis,* therapy, patient *outcome*, or society.

Effect: also effect size; see *Treatment effect*.

Effectiveness: the extent to which a specific intervention, procedure, regimen, or service does what it is intended to do under general conditions, rather than controlled conditions.

Efficacy: the extent to which a specific intervention, procedure, regimen, or service provide a beneficial result under controlled conditions.

Endpoint: *outcome* of interest.

Evidence-based approach: the systematic location and critical appraisal of published research and other available literature.

Evidence-Based Clinical Practice (EBCP): an emerging clinical discipline in which the best available evidence for research about *diagnosis*, prognosis, therapy, and other clinical and health issues is applied to decisions in health care.

Evidence table: a summary display of selected characteristics of studies of a particular issue of interest.

Experimental study: a type of epidemiological study design in which the exposure or intervention of interest is assigned to study subjects by the investigator often in a randomized manner (e.g., *randomized clinical trials*) to reduce *confounding*; in *evidence-based* terms, this type of study provides stronger evidence supporting a casual link between the intervention and *outcome(s)* of interest.

False negative: *"Type II"* or *"Beta"* error; a type of misclassification in which the disease is present but the test result is negative.

False positive: *"Type I"* or *"alpha"* error; a type of misclassification in which the disease is absent but the test result is positive.

FTEE: Full Time Equivalent Employee.

FY: Fiscal Year; VA's fiscal year begins October 1 and ends September 30.

Generalizability: the degree to which the inferences drawn from the study extend beyond the study sample; *external validity*.

Gold standard: reference test or criterion used to define the disease; the test to which the usefulness of the new test is compared.

Gray literature: "fugitive" literature; research reports not found in traditional peer-reviewed publications.

Health Service Research: the interdisciplinary study of the structures and processes through which personal health care services are organized, financed, delivered, and used.

HSR&D: Health Services Research & Development Service: a service within the Office of Research and Development, Veterans Health Administration, which examines how the organization, financing and management of health care affects treatment access, quality, cost, and outcome.

Historical control: group of study subjects who were not exposed to the variable of interest and who were observed at a different time period from the treatment group; the use of historical controls may affect the *internal validity* of a study.

Hypothesis testing: a means of interpreting the results of a clinical trial to determine whether an observed treatment effect could have occurred due to *chance* alone, given that a specified hypothesis were true; typically used to determine whether the *null hypothesis* can be rejected.

Incidence: the number of events (or *outcomes* of interest) occurring during a specified time period.

Kappa statistic: a measure of the degree of agreement that occurs between the diagnostic test and the gold standard over and above that which would have occurred by chance alone; can be used as a measure of test accuracy when there are more than two categories of test results.

Likelihood ratio: a method of expressing the diagnostic accuracy of complex imaging tests (or for revising the *pre-test probability*); the ratio of the probability of finding a particular image feature in patients with disease to the probability of finding the identical image feature in patients without the disease; this method allows for a comparison of the diagnostic value of various features.

Literature review: an overview or summary of research findings found in the literature; may range from unstructured and qualitative review to those more structured and systematic such as *meta-analyses*.

Management Decision and Research Center (MDRC): a program within *HSR&D Service* whose mission is to enhance the delivery of the highest quality health care by coupling the dynamic fields of health services research and management research and integrating these for managers and policymakers.

MDRC Technology Assessment Program: a program within the MDRC whose mission is to help VA researchers and managers make informed decisions about the acquisition and use of new medical technologies using an *evidence-based* approach.

Mean: measure of central tendency describing the average value of a group.

Median: the middle; measure of central tendency that divides a group into the lower half and upper half.

Medical technology: the drugs, devices, and medical and surgical procedures used in health care, and the organizational and supportive systems within which such care is delivered.

MEDLARS: Medical Literature Analysis and Retrieval System comprising about 40 computer databases managed by the National Library of Medicine.

MEDLINE: one of the most popular *MEDLARS* databases comprising bibliographic citations published since 1966 from about 3,700 health and biomedical journals.

MeSH: Medical Subject Headings; control vocabulary used in *MEDLARS* databases.

Meta-analyses: methods used to systematically identify, review, and statistically combine data from clinical studies to summarize the available evidence; particularly useful in summarizing prior research when individual studies are small, and they are individually too small to yield a valid conclusion.

Misclassification: the erroneous classification of an individual, a value, or an attribute into a category other than that to which it should be assigned.

Morbidity: any departure, subjective or objective, from a state of physiological or psychological well-being.

Mortality rate: the proportion of a population who die of a particular cause, usually expressed within a time interval of one year (i.e., death rate).

Moving target: term used to describe a technology that has rapidly changing properties.

Multiple regression: see *regression analysis*.

Negative Predictive Value: the proportion of those who test negatively who really do not have the disease.

Nonexperimental study: "observational study"; a type of epidemiological study that is based on existing exposure conditions without investigator intervention; this type of study is commonly used, but provides weaker evidence of a casual link between the intervention and outcome(s) of interest.

Null hypotheses: a statement used in *hypothesis testing* which says that the results observed in a study do not differ from what might have occurred as a result of chance alone; the intervention of interest has no effect upon the outcome studied.

Observational study: see *nonexperimental study*.

Odds: the ratio of the probability of occurrence of an event to that of nonoccurrence.

Outcome: the end result of health care that may stem from exposure to a casual factor, or from preventive or therapeutic interventions; may also include social and psychological function, patient attitude, health-related knowledge acquired by the patient, and health-related behavioral change.

Outlier: an observation differing so widely from the rest of the data as to lead one to suspect that a gross error may have been committed.

P value: a statement of the probability that the difference observed could have occurred by chance, reflecting the *statistical significance* of the result.

Patient selection bias: error due to systematic differences between those who are included in the study and those who are not; may affect *external validity* of a study.

Peer review (process): a process by which manuscripts are submitted to health, biomedical, other scientifically oriented journals, and other publications are evaluated by appropriate experts to determine whether the manuscript is of adequate quality for publication.

Positive Predictive Value: the proportion of those who test positively who really have the disease.

Post-test probability of disease: "posterior probability"; the probability of disease given the symptom.

Power: the probability of rejecting a *null hypothesis* when the *null hypothesis* is indeed false; the relative frequency with which a true difference of specified size between the comparison groups would be detected by the intervention or test of interest; expressed as 1- (the probability of a *Type II*) or (1-*Beta*).

Precision: the reproducibility of the study result, give similar circumstances, affected by patient and laboratory conditions, interobserver variation, and intraobserver variation.

Pre-test probability of disease: "prior probability"; the overall probability of disease among the population before knowing of the presence or absence of the symptom.

Prevalence: the number of instances of a given disease or occurrence in a given population at a specific point in time.

Prospective study: see *cohort study.*

PTF (Patient Treatment File): VA database that collects and maintains patient information, beneficiary classification and clinical information relative to diagnostic, surgical and treatment procedures.

Publication bias: an editorial preference for publishing particular findings, most notably studies demonstrating positive results over those which are negative.

Quality of life: a multidimensional construct denoting a wide range of capabilities, limitations, symptoms, and psychological characteristics that describe an individual's ability to function and derive satisfaction from a variety of roles.

Randomized Clinical Trial: an *experimental study* design in which eligible patients are randomly assigned to one or more treatment groups and a control group, and outcomes followed; the strongest of studies designed to establish causation.

ROC (Receiver Operating Characteristic) Curve: a graphic means for assessing the ability of a test to discriminate between diseased and nondiseased subjects; can be used to determine the optimal cut-off for a particular test or to compare accuracy of two diagnostic tests.

Registry: a system of ongoing registration for compiling data concerning all cases of a particular disease or other health-relevant conditions in a defined population such that the cases can be related to a population base.

Regression analysis: an approach that uses the best mathematical model (e.g., linear, logistic) to describe or predict the effect of an independent variable "X" on dependent variable "Y"; "multiple" regression involves estimating the effect of several independent variables on the dependent variable.

Relative Risk: a measure that describes the strength of the association between exposure and disease occurrence; the ratio of the occurrence of disease in the exposed to the occurrence of disease in the unexposed.

Reproducibility: see *precision.*

Resolution: the ability of an imaging device to distinguish two objects that are separate in either physical distance (spatial resolution) or in composition (contrast resolution).

Sample size: the total number of subjects in a study; including both treatment and control groups.

Sensitivity: the proportion of people with the disease who test positively.

Sensitive analysis: using a range of estimates of key variables in recalculations of a mathematical model or analysis to determine if changes in these estimates change the results of the analysis.

Series: see *case series.*

Specificity: the proportion of people without the disease who test negatively.

Staging: the classification of the severity of a disease in distinct stages on the basis of established signs and symptomatic criteria.

Statistical power: see *power*

Statistical significance: a conclusion determined by a *statistical test* that demonstrates whether a *technology* or intervention has a true effect on *outcome* over and above that which would have occurred by *chance*; it does not provide information about the magnitude of the effect, nor is it sufficient to demonstrate *clinical significance* of the *technology* or intervention on patient *outcome.*

Statistical test: a statistic (i.e., a mathematical formula or function) used to determine *statistical significance* by comparing the difference in outcomes of the comparison groups; examples are the F, t, Z, and chi-square tests.

Study base: the study population of interest observed over a specified period of time.

Systematic review: an overview prepared and appraised according to uniform, scientific principles, and which provide the highest level of evidence available; a *meta-analysis* is a type of systematic review which employs statistical methods for combining trials.

Technology: the drugs, devices, and medical and surgical procedures used in health care, and the organizational and supportive systems within which such care is delivered (Office of Technology Assessment, 1978).

Technology Assessment: any process of examine and reporting properties of a medical technology used in health care, such as safety, efficacy, feasibility, and indications for use, cost, and cost-effectiveness, as well as social, economic, and ethical consequences, whether intended or unintended; its purpose is to support technology-related policy making in health care.

Therapeutic impact: a characteristic of *diagnostic test efficacy* that describes the effect of a diagnostic test on therapeutic choices.

Type I error: *"Alpha"*; see *false-positive error*.

Type II error: *"Beta"*; see *false-negative error.*

Validity: the degree to which the inference drawn from a study sample extends beyond that study sample; based on *internal validity* (the index and comparison groups are selected and compared in such a manner that the observed differences between them may be attributed only to the exposure being studied) and *external validity* (generalizability of the results to a target population beyond the study population).

Appendix 3

Systematic Review:
PET as a Diagnostic Test in Head and Neck Cancer

Author: Karen Flynn, D.D.S., M.S., *Manager, MDRC Technology Assessment Program*

Appendix 3

Systematic Review: PET as a Diagnostic Test in Head and Neck Cancer

The final literature database searches for the systematic reviews were performed on September 10, 1996; the assessment represents peer-reviewed literature published and indexed as of that date.

This Appendix to the PET assessment presents the results of the systematic review of PET in head and neck cancer. A general rationale for the use of PET in oncology is supplied by Hawkins, et al. (1994) and Hoh, et al. (1994):

- many forms of cancer characteristically perturb tissue biochemical and physiological processes and PET imaging can be expected to detect the resulting abnormalities;
- reliance on tumor histology and anatomy limits the oncologist's tools for selecting optimal treatment;
- the ability to monitor metabolic responses to treatment could allow the early re-direction of therapy in patients who fail to respond to the first attempt at radiation or chemotherapy.

These and other authors (e.g., Price and Jones, 1995) report that PET studies in cancer are emerging as a major focus of the technology, both in basic research and in clinical investigations. Information gathered by the MDRC Technology Assessment Program from VA PET facilities corroborates that perception (see *Appendix 9: Experience With PET in VHA*).

Fluorine-18-fluorodeoxyglucose (FDG) is the most commonly employed radiopharmaceutical in PET cancer studies. Many neoplasms have high glycolytic rates, resulting in intracellularly trapped phosphorylated FDG that can be imaged with PET. Hawkins, et al. (1994), note that tumor-specific biochemical characteristics of glucose transport and phosphorylation may affect quantitative estimates of tumor glucose metabolism with FDG PET, and that investigations are under way to define these characteristics. However, these uncertainties may be of less concern with qualitative or semiquantitative FDG PET cancer studies because the primary intent of such studies is to detect and map tumor foci, not to rigorously quantify tumor glycolytic rates.

In some instances, PET imaging techniques have been modified to meet the needs of cancer diagnosis. Most PET systems allow axial fields of view (the length of the body encompassed by a series of cross sectional images) of approximately 10 cm. Cancer is frequently distributed beyond this field of view, and whole body image acquisition procedures have been developed (Hoh, et al., 1993). Since it is impractical to apply standard transmission scanning attenuation correction methods to these procedures, whole body PET imaging is primarily useful as a qualitative indicator of disease distribution.

Nieweg (1994) and Price and Jones (1995) define a number of potential applications for PET in oncology. These include:

- tumor detection (although PET images offer insufficient structural detail and should not be used to visualize anatomy; registration techniques to combine PET and anatomic imaging into a single image are under development to circumvent this limitation);
- staging (particularly using whole-body imaging methods) although there is a lower limit to the size of metastases that can be detected by PET;
- detection of local recurrence of disease, since anatomically-based imaging is often limited by the effects of treatment;
- prediction of tumor response to chemotherapy;
- treatment monitoring.

I. BACKGROUND

A. General sources

This overview is based on Vokes, et al. (1993), and on information distributed by the National Cancer Institute (NCI) through its on-line Physician Data Query (PDQ) system (accessed in September, 1996). Additional sources are cited in the text and included in the "References" section.

B. Description

This report will define head and neck cancer as the common squamous-cell carcinomas of the oral cavity, nasal cavity and paranasal sinuses, pharynx, and larynx. Skin, brain, ocular, thyroid, and salivary gland tumors and the rare tumors of other histopathologic types (sarcomas and lymphomas) that can have primary sites in the head and neck will not be discussed.

C. Epidemiology

Approximately 41,000 incident cases of head and neck cancer (3% of incident cases of all types of cancer), and 12,500 deaths (2% of all cancer related deaths) attributable to head and neck cancer are estimated for the United States in 1996 (American Cancer Society, 1996). Within the Veterans Health Administration, malignant neoplasms of the lip, oral cavity, and pharynx (but not larynx) accounted for a total of 3,361 hospital patients discharged (0.4% of all patients discharged within the system), with an average length of stay of 21.3 days, in fiscal year 1994 (Annual Report of the Secretary of Veterans Affairs, 1994).

Both incidence and mortality rates for head and neck cancer are substantially higher among men than among women (Spitz, 1994). Spitz (1994) reports that the incidence of oral and pharyngeal cancers decreased in white men of all ages from 1973 to 1989. However, there have been significant increases in incidence among black men; the incidence rates among black men is nearly double that of white men for those younger than 65 years, as are mortality rates.

The most important risk factors for all head and neck cancers are tobacco and excess alcohol use. Additional exposures that have been found to be associated with head and neck cancer include marijuana, occupational exposures (nickel refining, woodworking, the textile industry), and some viruses (particularly Epstein-Barr virus for nasopharyngeal cancer). Although sporadic cases of head and neck cancer occur in young adults and nonusers of tobacco and alcohol, most cases occur in males over 50 years of age with these risk factors.

D. Diagnosis

Mucosal surfaces in the upper aerodigestive tract, lungs, and esophagus are exposed to the same carcinogens, and multiple anatomic sites may be at risk for the simultaneous or sequential development of dysplastic and malignant lesions ("large field carcinogenesis"). Accordingly, there is a high incidence (in some reports as high as 25 to 30%) of synchronous (occurring at the same time) and metachronous (occurring later in time) second primary cancers in the head and neck, lung, and esophagus. Diagnostic, staging, and follow-up procedures are designed to monitor patients for second primaries as well as for the extent of the first tumor.

The signs and symptoms of head and neck cancer vary with the primary site and the stage of disease. Patients with early stage cancer may have only vague, nonspecific symptoms, and diagnosis requires a high index of suspicion among primary care physicians, oral surgeons, and general dentists. Higher stage disease is associated with increasingly severe symptoms at presentation. Some patients present with enlarged cervical lymph nodes but without an apparent primary mucosal surface tumor.

Initial diagnosis of head and neck cancer is based on physical examination, biopsy, indirect laryngoscopy or examination with a flexible fiberoptic nasopharyngoscope, radiologic evaluation (computed tomography or magnetic resonance imaging), and endoscopic examination (direct laryngoscopy, esophagoscopy, and bronchoscopy) with the patient under anaesthesia. Neither a liver-spleen scan nor a bone scan are generally felt to be of significant diagnostic value.

E. Staging, treatment, and survival

The TNM (tumor, node, metastasis, illustrated below) staging system integrates all clinical and imaging information, including the size of the primary lesion, involvement of adjacent structures and lymph nodes, and distant metastases. As in all malignancies, the stage of disease at diagnosis is a primary prognostic factor, with lower stage, locally confined disease associated with a higher probability of cure after treatment and longer survival than higher stage disease.

Table 1: Head and Neck Cancer TNM Staging System

	N_0	N_1	N_{2-3}	
T_1	I			Roman numerals represent stages
T_2	II			All M_1 tumors are stage IV
T_3		III		
T_4			IV	

T stages are separately defined for each anatomic region.
N_0-N_3 indicate a range of cervical lymph node involvement, from clinically negative to nodes > 6 cm.
M_0 indicates the absence of metastatic disease at distant sites; M_1 indicates its presence.

Approximately one-third of patients with head and neck cancer have lower stage, confined disease at diagnosis; most of the remaining patients have locally or regionally advanced disease (including spread to lymph nodes in the neck). Head and neck cancer which has already metastasized widely (e.g., to brain, lung, bone, or liver) at the time of presentation is less frequent. Standard therapy accordingly emphasizes local and regional approaches (surgery, radiation therapy, or both) with curative intent. With the exception of laryngeal cancer (for which induction chemotherapy may be used with radiation in an attempt to circumvent the need for laryngectomy), chemotherapy is generally accepted as standard therapy only for patients with recurrent or metastatic disease. In such patients, the intent of chemotherapy is palliative, rather than curative.

Table 2 Head and Neck Cancer: Stage of Disease, Standard Therapy, and Survival by Primary Site

Primary Site	Stage of Disease	Standard Therapy	5 Year Survival (3 year where noted)
All head and neck primary tumors	I II III or IV Inoperable, III or IV	Surgery or radiation Surgery or radiation Extensive surgery (including neck dissection) followed by radiation Radiation alone, sometimes followed by surgery	> 80% > 60% < 30% 2-25%
Nasal cavity/paranasal sinuses	I II III IV	Surgery and radiation Surgery and radiation Surgery and radiation Surgery and radiation	60-70% 60-70% 25-35% 10-25%
Nasopharynx	I II III IV	Radiation Radiation Radiation, followed by neck dissection as indicated Radiation; neck dissection for recurrent or persistent nodes	65-95% 50-65% 30-60% 5-50%
Oral cavity	I II III IV	Surgery or radiation, depending on anticipated functional result Surgery or radiation, depending on anticipated functional result Surgery and/or radiation, depending on site Surgery or radiation, depending on size and site of lesion	70-90% 50-80% 25-35% < 25%
Oropharynx	I II III IV	Surgery or radiation Surgery or radiation Surgery with post-operative radiation Surgery with post-operative radiation	60-100%, depending on site 50-100%, depending on site 20-30%, depending on site 14-20%, depending on site
Hypopharynx	I II III IV	Laryngopharyngectomy, occasionally with post-operative radiation Laryngopharyngectomy, occasionally with post-operative radiation Surgery with post-operative radiation Surgery with post-operative radiation	50-80% 50-60% 30-50% 15-25%
Larynx	I II III IV	Radiation Radiation Radiation; laryngectomy if persistent disease after radiation Total laryngectomy and followed by radiation	96-98% 80-94% 3-year, 45-75% 3-year, 10-35%
Metastatic squamous neck cancer with occult primary	 N1 N2 N3	Appropriate evaluation for primary in upper aerodigestive tract, esophagus, lung or genitourinary tract Radiation or neck dissection Radiation or neck dissection Radiation or neck dissection	3-year survival: 40-50% 25-30% 10-15%
Metastatic/recurrent disease, any primary site	Most treatment failures occur at site of original primary. Metastatic disease = IV	Surgery or radiation as feasible and dependant on first line treatment received Chemotherapy with palliative intent; further investigation into quality of life during chemotherapy needed.	Response lasts median 3-6 months; 40% of patients who receive combination chemotherapy alive at 9 months

Table 2 presents information on stage of disease at diagnosis, standard therapy, and survival. The NCI notes that there is a paucity of well-designed, controlled prospective studies comparing treatment modalities in patients with head and neck cancer, making it difficult to unequivocally state the ideal therapy for a specific site or stage of cancer originating in this anatomic area. The preferred treatment generally will depend on the skills of the treating physician, the needs of the patient, and a determination of the treatment which will cause the least functional disability. Ongoing clinical investigations for cancers at most sites in the head and neck focus on the addition of chemotherapy to surgery and radiation for local or regionally advanced disease in an attempt to reduce the need for surgical intervention and to improve cure and survival rates.

Since head and neck cancer is strongly associated with tobacco and alcohol use, many patients also have chronic heart, lung, and liver diseases. These comorbid conditions account for approximately 30% of deaths among patients with head and neck cancer. All of these tumors present complex medical, surgical, psychosocial, and rehabilitative problems, which frequently are managed by multidisciplinary groups including head and neck surgeons, radiation oncologists, medical oncologists, speech pathologists, nutritionists, dentists, pathologists, and diagnostic radiologists. Therapy for head and neck cancer inevitably has significant, sometimes profound, side effects, and even patients who are cured are often disfigured, lose their ability to speak and eat normally, and suffer the psychological morbidity associated with these disabilities.

F. Potential roles for PET

Diagnostic tests have an impact at several points in the initial work up and treatment of a patient with head and neck cancer. These include:

- initial diagnosis in the symptomatic patient, the patient with clinical signs of malignancy, or the patient with unexplained cervical lymphadenopathy [which occurs in 3% to 9% of patients with cancer of the head and neck (de Braud and Al-Sarraf, 1993)];

- decision making regarding specifics of treatment (in head and neck cancer patients a significant question is whether to enhance strictly local treatment at the primary site to include treatment of microscopic metastases to the cervical lymph nodes in a clinically negative neck);

- monitoring the results of treatment;

- post-treatment surveillance to define disease recurrence at the original primary site, or to define metastatic spread of disease.

Bailet, et al. (1992) note that CT and MRI have significantly improved the detection of occult cervical metastases in patients with head and neck cancer. Improved detection in turn has resulted in improved management of patients at high risk of cervical metastases (e.g., tumors of the base of the tongue, supraglottic larynx, and pyriform sinus). However, further improvements in the points at which imaging could impact patient management noted in the list above are still sought. Evaluation of head and neck tumors after surgery and/or radiation therapy can be complicated by the effects of treatment, making anatomically-based post-treatment imaging studies difficult to interpret (Chaiken, et al., 1993). Surgery inevitably results in deviations from normal anatomy, and radiation therapy can be associated with loss of tissue planes, edema, and residual masses. In this context, the information supplied by FDG PET on glucose metabolism in head and neck tumors could be clinically useful.

The results of FDG PET imaging in patients with head and neck cancer were first published by Bailet, et al., (1992) from the UCLA School of Medicine and its affiliated hospitals, including the West Los Angeles Veterans Administration Medical Center. This and subsequent studies indicated that FDG PET imaging of primary tumors and related cervical lymph node metastases and assessing tumor response to therapy was feasible. The table below summarizes the qualitative review by Mancuso, et al. (1994), of the initial experience at UCLA:

Potential Benefit of FDG PET	Currently Available Data Suggest...
identification of primary site in patients with cervical lymphadenopathy of unknown origin, allowing for more timelyfocused treatment of the primary. Second, synchronous primaries may be detected.	PET appears to be able to reliably detect primaries (including submucosal) of 1.0 cm and greater diameter; unpublished data indicate that PET can detect up to 50% of inapparent primaries, compared to 15-20% with CT or MRI.
Detection of subclinical cervical lymph node metastases, allowing more informed decisions re observation vs treatment in patients otherwise at low risk of cervical spread.	PET is limited in its ability to detect cervical lymph node metastases. Since FDG uptake is probably proportional to the number of cells in a metastatic lesion, microscopic nodal deposits are likely to produce false negative results. Reactive, metabolically hyperactive nodes may produce false positive results. No anatomic or physiologic study (including PET) is likely, in the near future, to detect microscopic disease accurately enough to make it the sole determinant of treatment of the neck.
Earlier detection of persistent or recurrent tumor, allowing more prompt salvage therapy.	It remains unknown whether PET will allow an earlier definition of treatment failure than CT or MRI. Further studies with longer follow up are needed, as are studies to define high-risk patients who would benefit from intensive post-treatment surveillance. Optimal post-treatment surveillance protocols remain undefined; no surveillance protocols have been demonstrated to be associated with improved survival.

Source: Mancuso, et al., 1994

II. RESULTS

Twenty-three articles were selected from the MEDLINE and other database searches and from the bibliographies of initially retrieved articles as meeting the screening criteria. After review, 9 (39%) were found to meet the criteria for assignment to the following levels of the diagnostic efficacy hierarchy (Fryback and Thornbury, 1991; *Appendix 2: Assessing Diagnostic Technologies*): 4 met the definition of technical efficacy; 4 met some of the evidence-based criteria for diagnostic test evaluations (Table 4), and one additional study met some of the evidence-based criteria while comparing PET to MRI and made an attempt to extrapolate findings to therapeutic efficacy (Table 5). Table 3 summarizes cross-study findings on PET and alternative technologies.

All currently available data on the use of PET in patients with head and neck cancer are based on case series studies, which provide Level V (i.e., the weakest) evidence of any association between the use of a technology and improved patient outcomes. Some studies, however, have internal controls for subsets of head and neck cancer patients (e.g., patients with and without cervical node involvement).

Studies classified at the "technical efficacy" level of the diagnostic efficacy hierarchy (Fryback and Thornbury,1991) are listed in Section VII, below. The definition of technical efficacy was expanded to include studies that were not designed to assess diagnostic accuracy or that did not meet the evidence-based criteria for diagnostic accuracy. These studies did provide information necessary to subsequent diagnostic efficacy studies. Data abstraction tables for technical efficacy studies are on file with the MDRC Technology Assessment Program.

Table 4 abstracts data from the studies that assessed the diagnostic accuracy of PET for certain applications in head and neck cancer; these studies also compared PET directly to other imaging technologies. The diagnostic accuracy data reported in Table 4 apply only to detection of cervical lymph nodes and distinguishing recurrent disease from treatment artifacts. The studies in Table 4 did not include control groups without head and neck cancer or with diseases that need to be distinguished from head and neck cancer, and accordingly did not meet evidence-based medicine criteria for diagnosing primary disease. Table 5 abstracts data from the one retrospective, hypothetical therapeutic efficacy study. The MDRC Technology Assessment Program was unable to locate any studies using PET in head and neck cancer at the patient outcome or societal levels of the diagnostic efficacy hierarchy.

Methodologic and sample size limitations of the studies of PET in head and neck cancer argue for caution in interpreting the sensitivity and specificity reported in Tables 4 and 5. Only one of the studies in Table 4 (Lapela, et al., 1995) blinded image interpreters. It was decided that meta analyses of the diagnostic accuracy studies would not yield further insights into PET's usefulness as a diagnostic test, due to the potential for significant bias in the design of these studies. Qualitative results, organized by the potential role of PET in the management of head and neck cancer, are:

A. Detecting unknown primaries in patients who present with metastatic cervical nodes

The MDRC Technology Assessment Program was unable to locate any PET studies that met evidence-based criteria for diagnosis of unknown primaries.

B. Detecting primary disease

The MDRC Technology Assessment Program was unable to locate any PET studies that met evidence-based medicine criteria for diagnosis of primary disease.

C. Detecting cervical metastases

A number of studies partially met evidence-based medicine criteria for diagnostic test evaluations. One study (Benchaou, et al., 1996) met all evidence-based criteria and received a good methodologic quality score. These studies suggest that PET is somewhat limited in its ability to detect subclinical cervical node metastases: a high rate of false positives for cervical nodes is associated with PET in the studies in Table 4, and is attributed to the metabolic activity of reactive lymph nodes. The available evidence suggests that PET does not perform substantially better in this setting than do MRI, CT, or ultrasound-guided fine needle aspiration biopsy.

One study (Braams, et al., 1995; Table 5) was classified at the therapeutic efficacy level (Level 4, detailed in *Appendix 2: Assessing Diagnostic Technologies*), because the authors extrapolated diagnostic accuracy to a retrospective, hypothetical decision regarding performing neck dissection in their small series of patients. While this study indicates that

PET may impact clinical management and patient outcomes, its small size and hypothetical nature suggest that further documentation would be needed to define marginal benefits over anatomic imaging.

D. Detecting recurrent disease

Lapela, et al. (1995), found that blinded visual interpretation of PET and blinded interpretation of CT had approximately equivalent sensitivity and specificity in detecting recurrent disease. An unblinded study (Rege, et al., 1994) found that PET was superior to MRI in detecting recurrent disease.

III. SUMMARY

Table 3 summarizes published findings on the diagnostic accuracy efficacy of PET and its alternatives in diagnosing cervical lymph node involvement with disease and in evaluating suspected disease recurrence. Only one study (Lapela, et al., 1995) met all evidence-based medicine criteria for diagnostic test evaluations; the unit of analysis in this study was regions, not patients. While data on other uses of PET obtained in uncontrolled studies are also included in Table 3, the MDRC Technology Assessment Program was unable to locate any published studies that met completely evidence-based medicine criteria for evaluations of diagnostic tests for the use of PET in these settings. PET and CT have been compared in one retrospective, hypothetical therapeutic efficacy study, which also supplies diagnostic accuracy information (Braams, et al., 1995). The results of studies that did not blind image interpreters to disease status should be interpreted with caution. All of the studies listed in Table 3 received low methodologic quality grades due to the absence of blinding, the absence of controls, and/or small sample sizes.

IV. DISCUSSION

Authors of studies intended to document diagnostic accuracy generally concede that PET supplies information that can be complementary to, but that does not replace, anatomic imaging information in the management of head and neck cancer patients. In clinical use, PET would be incorporated into a diagnostic test battery, and information on pre- and post-test probabilities of disease at each step in the diagnostic strategy would be needed to define the marginal information yield associated with each of the tests (including PET).

Any analysis of the effect of PET on the outcomes of treatment which might be attempted, based on longer follow up of patients who have been reported in the existing literature, would be further complicated by the wide range of primary sites and stages of squamous cell cancer of the head and neck included in the case series, and the associated, correspondingly wide range of site-specific treatments and outcomes.

Mancuso, et al. (1994), note that FDG PET must be cost competitive with CT and MRI and/or offer a significant increment of improvement in detection if its use is to be justified. Other competing technologies may be under development. Drane, et al. (1994), report a technique which combines FDG with SPECT, and which may be associated with a lower per scan cost than PET and with wider availability. Gamma cameras are under development which permit imaging of 511-keV photons from positron emitters such as FDG. Eighteen patients with head and neck tumors were included by Drane, et al. (1994) in an initial study. However, the reporting of results in the published report was meager; the authors caution that the study was intended only to support the feasibility of such imaging, not to determine the diagnostic accuracy of FDG SPECT.

Alternative imaging protocols have been proposed in an effort to reduce the number of unnecessary neck treatments in patients with head and neck cancer. Baatenburg de Jong, et al. (1993), report the results of a diagnostic thinking efficacy study. Ultrasonography has a high sensitivity (97%) for detection of metastatic involvement of the neck. The specificity is low (32%) unless it is combined with ultrasound-guided fine needle aspiration biopsy; Braatenburg de Jong, et al., found a specificity of 93% for the combined technique. These authors used pre-test probabilities of disease according to anatomic site (from the literature) and a range of sensitivities and specificities to calculate post-ultrasound-guided fine needle aspiration biopsy probability of disease. Clinicians could apply these results to the treatment threshold probabilities in use at their institutions (e.g. at some institutions all patients with a probability of occult metastases > 5% receive elective neck dissections).

Weiss, et al., (1994) provide guidance concerning treatment thresholds. These authors used decision analysis to plan management for patients with head and neck cancer and clinically negative necks, using clinical staging information and the probabilities of occult cervical metastases associated with each stage. Their objective was to generate an optimal threshold (for the probability of occult cervical metastases) beyond which treatment would be given. Based on their analysis, these authors found that it is reasonable to observe patients with a probability of occult metastases less than 20%, while treatment is warranted in the presence of a probability greater than 20%.

V. SUGGESTIONS FOR FURTHER RESEARCH

The types of study designs, and the strength of the resulting evidence from further research into the role of PET in head and neck cancer care, will be inherently constrained by a number of factors. The epidemiologic data cited earlier in the discussion of head and neck cancers indicate that these cancers are relatively rare, and collecting enough cases of such cancers for some study designs (e.g., prospective or cohort studies) may be difficult.

1) PET has potential uses at several points in the diagnosis and management of head and neck cancer patients. An early step in defining these uses is determination of diagnostic accuracy. Studies that have been published to date generally have methodologic weaknesses, and may overestimate accuracy. Controlled, blinded studies should be conducted; multi-center studies may be needed to accrue meaningful numbers of patients.

2) The role of PET in modifying treatment decisions or improving the outcomes of head and neck cancer therapy is currently limited to one retrospective, hypothetical study with significant methodologic limitations. Prospective studies should be conducted, and again may need to involve multiple centers to accrue meaningful numbers of patients.

3) A VA PET registry could provide a range of data on demographic and clinical characteristics of patients on whom PET studies are performed, and on their clinical outcomes in a variety of settings; While a registry would not provide the strength of evidence associating PET would improved outcomes that would be provided by randomized clinical trials, it would circumvent the problem of low disease prevalence.

4) The role of PET as part of a diagnostic test battery should be defined.

Table 3 **Summary of the literature:** **Diagnostic accuracy efficacy of PET and alternatives in head and neck cancer**
(from studies comparing PET directly to other diagnostic tests)

Role	Study	N	Operating characteristics*				Evidence-based medicine criteria**			Methodologic quality grade***
			PET	*CT*	*MRI*	*Other*	*controls*	*standard*	*blinding*	
Unknown primary	Rege, et al., 1994	4 cases 0 controls	Se = 50%		Se = 0%		-	+	-	D
Known primary site	Rege, et al., 1994	30 cases 0 controls	Se = 97%		Se = 77%		-	+	-	D
	Laubenbacher, et al., 1995	17 cases 0 controls	Se = 100%		Se = 100%	endoscopy, Se = 100%	-	+	-	D
Primary tumor staging (size, extent)	Laubenbacher, et al., 1995	17 cases 0 controls	Se = 41%		Se = 41%	endoscopy, Se = 59%	-	+	-	D
Cervical node involvement	Rege, et al., 1994	16 cases 18 controls	Se = 88% Sp = 89%		Se = 81% Sp = 89%		+	+	-	D
	McGuirt, et al., 1995	14 cases 31 controls	accuracy = 82%	accuracy = 82%		clinical exam accuracy = 71%	+	+	-	D
	Laubenbacher, et al., 1995	83 pos nodes 438 neg nodes	Se = 90% Sp = 96%		Se = 78% Sp = 71%		+	+	-	D
		18 pos neck sides 16 neg neck sides	Se = 89% Sp = 100%		Se = 72% Sp = 56%		+	+	-	D
	Braams, et al., 1995	22 pos nodes 177 neg nodes	Se = 91% Sp = 88%		Se = 36% Sp = 94%		+	+	-	D
	Benchaou, et al., 1996	54 pos node groups 414 neg node groups	Se = 72% Sp = 99% PPV = 89% NPV = 99%	Se = 67% Sp = 97% PPV = 74% NPV = 95%		clinical exam Se = 61% Sp = 97% PPV = 72% NPV = 95%	+	+	+	B
Suspected recurrent disease	Rege, et al., 1994	10 cases 7 controls	Se = 90% Sp = 100%		Se = 67% Sp = 57%		+	+	-	D
	Lapela, et al., 1995	16 pos 17 neg	Se = 88 -94% Sp = 43 -86% depending on criteria for pos	Se = 92% Sp = 50%			+	+	+	C

Abbreviations:
Ct, computed tomography
MRI, magnetic resonance imaging
neg, negative for disease
pos, positive for disease
Se, sensitivity
Sp, specificity
PPV, positive predictive value
NPV, negative predictive value
US/FNA, ultrasound/fine needle aspiration

* operating characteristics defined in Appendix 2: Assessing Diagnostic Technologies, pages 5-7
** Appendix 2, page 8
*** Appendix 2, page 9

Table 4 Diagnostic accuracy efficacy of FDG PET and anatomic imaging in detecting cervical lymph nodes from head and neck cancer

Notes *All of the studies in the table are cases series (Level V evidence); most of the studies did not meet evidence-based medicine criteria for evaluations of diagnostic tests for primary head and neck cancer (because no patients without head and neck cancer were included). However, there were internal controls for subsets of patients (e.g. those with cervical lymph nodes positive for disease), and it was possible to calculate sensitivity and specificity for PET in those subsets.*

Some of the studies in the table also do not meet the evidence-based medicine requirement for blinding; sensitivity and specificity reported in these studies should be interpreted with caution.

Where substantial duplication in purpose of study, patients studied, and results in multiple studies from the same institution could be inferred, only the most recent, largest, most rigorously designed, or most comprehensive was included in the table. Studies reviewed but not included are listed in Section VIII.

The reference test for the PET operating characteristics reported in the "Results/Comments " column is biopsy.

Study	***Patients/Methods***	***Results/Comments***
Rege, et al., 1994 *UCLA School of Medicine, West LA VAMC*	**Purpose** • to summarize cumulative (3 year) experience with PET as a supplement to anatomic imaging in H&N tumors • PET and MRI compared to histopathology **Cases** 60 patients with biopsy-proven H&N cancers (53 SCC, 7 other types) recruited after presenting for evaluation of H&N tumors: • 34 patients scanned before treatment (staging); 15 of these received serial scans to monitor response to treatment; 4 had unknown primary tumors and metastatic cervical nodes • 19 patients evaluated for recurrent disease • 7 patients with advanced disease receiving palliative laser therapy (not included in analyses) **Methods** • all patients received MRI and PET • MRI interpreted by 2 experts with full clinical information available • PET interpreted visually by consensus of 3 nuclear medicine experts • FDG uptake in tumor/nodes compared to cerebellum • MRI used as anatomic template to locate increased FDG uptake seen on PET **Limitations of study design** • Images interpreted without blinding • blood glucose status of patients not noted	**Staging/cervical nodes (16 cases with + nodes, 18 cases with - nodes):** • PET: *Se = 87.5%; *Sp = 89%; *PPV = 87.5%; *NPV = 89% • MRI: *Se = 81%; *Sp = 89%; *PPV = 87%; *NPV = 84% **Patients evaluated for recurrent disease (10 cases confirmed by histopathology; 7 cases negative histopathology for PET, 6 for MRI):** • PET: *Se = 90%; *Sp = 100%; *PPV = 100%; *NPV = 87.5% • MRI: *Se = 66.7%; *Sp = 57%; *PPV = 66.7%; *NPV = 57% **Cervical metastatic nodes/unknown primary (4 cases, no controls):** • PET identified primary disease in 2/4 patients • MRI negative in 4/4 **Detecting known primary (30 cases, no controls):** • PET detected known primary in 29/30 patients • MRI detected known primary in 23/30 patients **Treatment monitoring:** some patients experience transient increase in FDG uptake during treatment **General findings:** • PET supplies information which is complementary to, but does not replace, MRI and CT anatomical information in the management of H&N cancer patients • reactive lymph nodes/inflammation may cause false-positive PET results

Study	*Patients/Methods*	*Results/Comments*
McGuirt, et al., 1995 *Bowman Gray School of Medicine*	**Purpose** • to examine ability of PET to detect metastatic tumor in cervical nodes • to compare PET and standard diagnostic methods **Cases** 45 patients with variety of SCC seen at H&N tumor clinic • status of necks after dissection: 70% N0, 19% N1, 11% N2 **Methods** • visual interpretation of PET and calculation of SUV for lymph nodes showing increased FDG uptake • CT, clinical exam, and PET results compared to histopathology **Study design limitations** blinding of diagnostic test interpreters not noted	**Test characteristics: cervical node involvement** • PET: *Se = 83%; *Sp = 82%; accuracy = 82% • CT: (insufficient information provided to calculate Se, Sp, PPV, NPV); accuracy = 82% • clinical exam: (insufficient information provided to calculate Se, Sp, PPV, NPV); accuracy = 71% **Agreement of PET and CT** • in 84% of cases • where did not agree, CT more often correct re pathology but PET helped to clarify equivocal CT results **Authors' comments** • CT had lower false negative results than reported in literature; attributed to advances in CT imaging and resolution • FDG use in PET will always be associated with relatively high false-positive rates due to metabolic activity of reactive nodes • PET does not provide anatomic information necessary to surgeons • PET helpful when CT equivocal but cost and information yield of PET compared to CT argue against routine clinical use of PET
Laubenbacher, et al., 1995 *Technical University of Munich, Germany*	**Purpose** • to assess correlation between FDG uptake in primary tumors and histologic grade • to assess contribution of FDG PET to diagnostic accuracy in preoperative assessment of primary tumor and lymph node status • to determine whether attenuation correction is necessary for detection and staging of tumors with PET **Cases** 22 consecutive patients referred for surgery for SSC • range of stages, node status, grade in 17 patients at surgery • histopathologic confirmation in 5 patients with inoperable tumors not obtained **Methods** • whole-body PET performed after overnight fast; blood glucose levels recorded • qualitative (implied) and quantitative analysis of PET images • all patients had MRI, endoscopy, and histopathologic diagnosis/grading of tumors **Study design limitations** • blinding not noted for qualitative image analysis • surgeons had access to PET information and completeness of cervical dissections not reported (work-up bias?) • size criteria only were used for MRI diagnosis of node involvement, which may have decreased Se and Sp compared to diagnosis using additional criteria (e.g. central inhomogeneity)	**FDG uptake** • all primary tumors visualized on PET • no statistically significant difference in SUVs for primary tumor (range, 2.0 - 13.8) vs lymph nodes (range, 1.4 - 11.4) • no plateau in FDG uptake within 60 minutes post-injection • no significant correlation between FDG uptake and blood glucose levels, lesion size, or histologic grade • no significant differences between attenuation-corrected and noncorrected images for staging **T staging** • all tumors clearly visualized with PET, MRI, endoscopy • best results in endoscopy (correct staging in 10/17 cases) • MRI and PET each correctly staged 7/17 cases and overstaged ≈ 50% of cases **N staging** • 521 nodes assessed at surgery in 17 patients (34 neck sides) - 83 positive, 438 negative for metastases - 18 neck sides positive, 16 negative • individual node analysis: - PET: Se = 90%; Sp = 96%; PPV = 80%; NPV = 98% - MRI: Se = 78%; Sp = 71%; PPV = 34%; NPV = 95% • neck side analysis: - PET: Se = 89%; Sp = 100%; PPV = 100%; NPV = 95% - MRI: Se = 72%; Sp = 56%; PPV = 65%; NPV = 64% **Authors' comments** abbreviated protocol with emission scans without attenuation correction appears to fulfill clinical requirements

Study	Patients/Methods	Results/Comments
Lapela, et al., 1995 *University of Turku, Finland*	**Purpose** • to estimate quantitative FDG uptake values that would suggest the recurrence of head and neck cancer strongly enough to justify surgery • to compare visual, static, and kinetic analyses of FDG uptake in differentiating malignant and benign lesion with patients with previously treated H&N cancer **Cases** 15 patients who presented to otolaryngology department for evaluation of suspected recurrence of SCC after surgery and/or radiation therapy • 2/15 received second PET study for second recurrence **Methods** • blinded, independent visual analysis of 17 PET images by 3 investigators • PET images graded as clearly malignant, suspect, or negative • ROIs defined and SUVs calculated • SUVs used in static and kinetic analyses • CT performed in 13/15 patients and interpreted by one blinded radiologist	**Histopathology** 25 regions in 17 PET studies: 16 malignant, 7 not malignant (2 left out of analyses because PET and histology negative at time of study but recurrence documented within 6 months) **Interobserver variation** • agreement among all 3 PET readers for 20/25 regions • 2/3 readers agreed on remaining 5/25 regions **PET characteristics: blinded visual interpretation** • malignant + suspect lesions = positive: Se = 94%; Sp = 43% • malignant only = positive: Se = 88%; Sp = 86% **Quantitative analysis of PET** • median SUV of benign and malignant lesions significantly different ($p = .008$) • median regional metabolic rates also significantly different ($p = .002$) • at SUV = 5.74 cut-point, Se = 75% • at metabolic rate = 15.4 µmol/100g/min cut-point, Se = 86% **CT characteristics: blinded interpretation** •18 regions analyzed • Se = 92%; Sp = 50% **Authors' comments** • complex static or kinetic modelling provides no clear advantage over SUV or regional metabolic rates • SUVs at different institutions may not be directly comparable • quantitative analysis of small lesions (compared to PET resolution) should be performed with caution • PET false-positives attributed to reactive nodes
Benchaou, et al., 1996 *Geneva University Hospital, Switzerland*	**Purpose** to compare the results of PET, CT, and cervical node palpation in N-staging prior to surgery **Cases** 40 SCC; surgery indicated in all patients • 38 primary SSC, stages T1-3 • 2 unknown primary **Controls** 8 with benign or other tumors; surgery indicated in all patients • 6 benign neck masses • 2 cervical lymphoma **Methods** • all patients received clinical exam, CT, PET, endoscopy, surgery with neck dissection (9 node groups/neck side; 4 bilateral and 44 unilateral dissections) and histopathologic confirmation of tumor type and node status • blinded reading (PET qualitative/semi quantitative with lymphoid tissue as reference) of each type of test by certified specialist	**Node status** • 468 node groups examined (54 positive, 414 negative) • 23 patients N0 (15 SSC, 8 other/benign) • 25 patients N1 or N2 **Operating characteristics of tests** • PET: Se = 72%; Sp = 99%; accuracy = 96%; PPV = 89%; NPV = 99% • CT: Se = 67%; Sp = 97%; accuracy = 93%; PPV = 74%; NPV = 95% • Palpation: Se = 61%; Sp = 97%; accuracy = 93%; PPV = 72%; NPV = 95% **Statistical tests** • Se of PET significantly higher than Se of palpation ($p = 0.03$) • Se of PET equivalent to Se of CT ($p = 0.25$) • Sp of all tests equivalent ($p = 1$) • PPV of PET significantly higher than palpation ($p < 0.05$)

Abbreviations:
SCC, squamous cell carcinoma
H&N, head and neck
Se, sensitivity
Sp, specificity
PPV, positive predictive value
NPV, negative predictive value
RT, radiation therapy
ROI, region of interest
SUV, standardized uptake value = (tissue activity x weight)/ injected dose

* indicates calculated by MDRC TA Program from data supplied in published article

Table 5 **Therapeutic efficacy of FDG PET and anatomic imaging in detecting cervical lymph nodes positive for cancer and potential impact on decision to perform neck dissection**

Notes *The study in the table does not meet the evidence-based medicine requirement for blinding; sensitivity and specificity should be interpreted with caution.*

Case series (Level V evidence).

Study	*Patients/Methods*	*Results/Comments*
Braams, et al., 1995 *University Hospital Groningen, The Netherlands*	**Purpose** to investigate usefulness of PET in identifying lymph node metastases, compared to clinical and MRI findings **Cases** 12 patients presenting to H&N oncology group for evaluation of SCC at variety of primary sites **Methods** • whole body PET, MRI, clinical palpation for nodes, and histologic confirmation of node status for all patients during radical or modified radical neck dissection • PET images analyzed visually by two observers at same time • ROIs drawn and SUVs calculated • characteristics of tests calculated • false negative and false positive rates for PET and MRI and institutional criteria for elective and obligatory neck dissections retrospectively applied to hypothetical decision in study subjects **Study design limitations** blinding of image interpreters not noted	**Histopathology of resected specimens** • 22 metastatic lymph nodes • 25 reactive nodes •152 normal nodes **Characteristics of tests** • PET: Se = 91%; Sp = 88%; PPV = 48%; NPV = 99%; false positive rate = 52% • MRI: Se = 36%; Sp = 94%; PPV = 44%; NPV = 92%; false positive rate = 55% **SUV analysis** • metastatic node SUV = 2.5 ± 0.8 • reactive node SUV = 2.6 ± 1.4 (difference between reactive and metastatic nodes not significant) • normal node SUV = 1.0 ± 0.3 (difference between normal and metastatic/reactive nodes significant, $p < 0.001$) **Therapeutic efficacy: perform neck dissection?** • PET: dissection would have been performed in all patients with metastatic disease, and 5 patients would have received unnecessary dissections (false positives) (79% correct decisions) • MRI: 4 patients with metastatic disease would not have received dissections (false negatives), and 4 patients would have received unnecessary dissections (false positives) (66% correct decisions) **Additional findings and authors' comments** • PET false negatives attributed to small size of nodes (< 4 mm) or low SUV and partial volume effect • MRI false negatives in nodes < 10 mm • problem with PET is distinguishing metastatic and reactive nodes (high false positive rate); SUVs provided no additional diagnostic value over visual analysis; additional method to improve false positive rate desirable • sensitivity and specificity depend on number of lymph nodes retrieved during dissections • importance of PET is that no false-negative decisions re neck dissection would have been made

Abbreviations:
SCC, squamous cell carcinoma
H&N, head and neck
Se, sensitivity
Sp, specificity
PPV, positive predictive value
NPV, negative predictive value
RT, radiation therapy
ROI, region of interest
SUV, standardized uptake value

* indicates calculated by MDRC TA Program from data supplied in published article

VI. REFERENCES Background and diagnostic accuracy efficacy studies

American Cancer Society. *Cancer Facts & Figures - 1996.*

Baatenburg de Jong RJ, Rongen RJ, Verwoerd CDA, van Overhagen H, Lameri JS, Knegt P. Ultrasound-guided fine-needle aspiration biopsy of neck nodes. *Archives of Otolaryngology and Head and Neck Surgery* 1991;117:402-4.

Baatenburg de Jong RJ, Knegt P, Verwoerd CDA. Reduction of the number of neck treatments in patients with head and neck cancer.*Cancer* 1993;71:2312-8.

Beenken SW, Maddox WA, Urist MM.Workup of a patient with a mass in the neck. *Advances in Surgery* 1995;28:371-83.

Benchaou M, Lehmann W, Slosman DO, Becker M, Lemoine R, et al. The role of FDG-PET in the preoperative assessment of N-staging in head and neck cancer. *Acta Otolaryngologica* 1996;116:332-5.

Braams JW, Pruim J, Freling NJM, Nikkels PGJ, Roodenburg JLN, Boering G, et al. Detection of lymph node metastases of squamous cell cancer of the head and neck with FDG-PET and MRI. *Journal of Nuclear Medicine* 1995;36:211-6.

de Braud F, Al-Sarraf M. Diagnosis and management of squamous cell carcinoma of unknown primary tumor site of the neck. *Seminars in Oncology* 1993;20:273-8.

van den Brekel MWM, Castelijns JA, Stel HV, Luth WJ, Valk J, van der Waal I, Snow GB. Occult metastatic neck disease: detection with US and US-guided fine needle aspiration cytology. *Radiology* 1991;180:457-61.

Drane WE, Abbott FD, Nicole MW, Mastin ST, Kuperus JH. Technology for FDG SPECT with a relatively inexpensive gamma camera. Work in progress. *Radiology* 1994;191:461-5.

Lapela M, Grenman R, Kurki T, Joensuu,H, Leskinen S, Lindholm P, et al. Head and neck cancer: detection of recurrence with PET and 2-[F-18]fluoro-2-deoxy-D-glucose. *Radiology* 1995; 197:205-11.

Laubenbacher C, Saumweber D, Wagner-Manslau C, Kau RJ, Herz M, Avril N, et al. Comparison of fluorine-18-fluorodeoxyglucose PET, MRI and endoscopy for staging head and neck squamous-cell carcinomas. Journal of Nuclear Medicine 1995;36:1747-57.

Mancuso AA, Drane WE, Mukherji SK. The promise of FDG in diagnosis and surveillance of head and neck cancer. *Cancer* 1994;74:1193-5.

McGuirt WF, Williams DW, Keyes JW, Greven KM, Watson NE, Geisinger KR, Cappellari, JO. A comparative diagnostic study of head and heck nodal metastases using positron emission tomography. *Laryngoscope* 1995;105:373-7.

Minn H, Joensuu H, Ahonen A, Klemi P. Fluorodeoxyglucose imaging: a method to assess the proliferative activity of human cancer in civo. Comparison with DNA flow cytometry in head and neck tumors. *Cancer* 1988;61:1776-81.

Price P, Jones T, for EC PET Oncology Concerted Action and EORTC PET Study Group. Can positron emission tomography (PET) be used to detect subclinical response to cancer therapy? *European Journal of Cancer* 1995;12:1924-7.

Rege SD, Chaiken L, Hoh CK, Choi Y, Lufkin R, Anzai Y, et al. Change induced by radiation therapy in FDG uptake in normal and malignant structures of the head and neck: quantitation with PET. *Radiology* 1993;189:807-12.

Rege S, Maass A, Chaiken L, Hoh CK, Choi Y, Lufkin R, et al. Use of positron emission tomography with fluorodeoxyglucose in patients with extracranial head and neck cancers. *Cancer* 1994;73:3047-58.

Spitz MR. Epidemiology and risk factors for head and neck cancer. *Seminars in Oncology* 1994; 21:281-8.

Vokes EE, Weichselbaum RR, Lippman SM, Hong WK. Medical progress: head and neck cancer. *New England Journal of Medicine* 1993;328:184-94,

Weiss MH, Harrison LB, Isaacs RS. Use of decision analysis in planning a management strategy for the stage NO neck. *Archives of Otolaryngology and Head and Neck Surgery* 1994;120:699-702.

VII.REFERENCES: Technical efficacy studies

Greven KM, Williams DW, Keyes JW, McGuirt WF, Watson NE, Randall ME, et al. Positron emission tomography of patients with head and neck carcinoma before and after high dose irradiation. *Cancer* 1994;74:1355-9.

Haberkorn U, Strauss LG, Haag D, Dimitrakopoulou A, Ziegler S, Oberdorfer F, et al. Glucose uptake, perfusion, and cell proliferation in head and neck tumors: relation of positron emission tomography to flow cytometry. *Journal of Nuclear Medicine* 1991;32:1548-55.

Reisser C, Haberkorn U, Dimitrakopoulou-Strauss A, Seifert E, Strauss LG. Chemotherapeutic management of head and neck malignancies with positron emission tomography. *Archives of Otolaryngology and Head and Neck Surgery* 1995;121:272-6.

Wong WL, Chevretton E, McGurk M., Croft, D. PET-FDG imaging in the clinical evaluation of head and neck cancer. *Journal of the Royal Society of Medicine* 1995;88:469p-73p.

VIII. REFERENCES: Excluded studies

Exclusion criteria included:

- number of cases < 12
- duplicated or superseded by subsequent or concurrent study from the same institution
- radiopharmaceutical other than FDG
- gamma camera rather than PET
- insufficient information to judge comparability of case and control groups, details of imaging protocol, whether visual or quantitative analysis of PET data used, or type of PET data analysis used

Anzai Y, Carroll WR, Quint DJ, Bradford CR, Minoshima S, et al. Recurrence of head and neck cancer after surgery or irradiation: prospective comparison of 2-deoxy-2-[F-18]fluoro-D-glucose PET and MR imaging diagnoses. *Radiology* 1996;200:135-41.

Austin JR, Wong FC, Kim EE. Positron emission tomography in the detection of residual laryngeal carcinoma. *Otolaryngology and Head and Neck Surgery* 1995;113:404-7.

Bailet JW, Abemayor E, Jabour BA, Hawkins RA, Ho C, Ward PH. Positron emission tomography: a new, precise imaging modality for detection of primary head and neck tumors and assessment of cervical adenopathy. *Laryngoscope* 1992;102:281-8.

Bailet JW, Sercarz JA, Abemayor E, Anzai Y, Lufkin RB, Hoh CK. The use of positron emission tomography for early detection of recurrent head and neck squamous cell carcinoma in postradiotherapy patients. *Laryngoscope* 1995;105:135-9.

Berlangieri SU, Brizel DM, Scher RL, Schifter T, Hawk TC, Hamblen S, et al. Pilot study of positron emission tomography in patients with advanced head and neck cancer receiving radiotherapy and chemotherapy. *Head and Neck* 1994:340-6.

Chaiken L, Rege S, Hoh C, Choi Y, Jabour B, Juillard G, et al. Positron emission tomography with fluorodeoxyglucose to evaluate tumor response and control after radiation therapy. *International Journal of Radiation Oncology Biology and Physics* 1993;27:455-64.

Haberkorn U, Strauss LG, Dimitrakopoulou A, Seiffert E, Oberdorfer F, Ziegler S, et al. Fluorodeoxyglucose imaging of advanced head and neck cancer after chemotherapy. *Journal of Nuclear Medicine* 1993;34:12-7.

Jabour BA, Choi Y, Hoh CK, Rege SD, Soong JC, Lufkin RB, et al. Extracranial head and neck: PET imaging with 2-[F-18]fluoro-2-deoxy-D-glucose and MR imaging correlation. *Radiology* 1993; 186:27-35.

Lindholm P, Minn H, Leskinen-Kallio S, Bergman J, Ruotsalainen U, Joensuu, H. Influence of the blood glucose concentration on FDG uptake in cancer - a PET study. *Journal of Nuclear Medicine* 1993;34:1-6.

McGuirt WF, Greven KM, Keyes JW, Williams DW, Watson NE, Geisinger KR, Cappellari JO. Positron emission tomography in the evaluation of laryngeal carcinoma. *Annals of Otology Rhinology and Laryngology* 1995;104:274-8.

Minn H, Joensuu H, Ahonen A. Klemi P. Fluorodeoxyglucose imaging: a method to assess the proliferative activity of human cancer in vivo. Comparison with DNA flow cytometry in head and neck tumors. *Cancer* 1988; 61:1776-81.

Minn H, Paul R, Ahonen A. Evaluation of treatment response to radiotherapy in head and neck cancer with fluorine-18 fluorodeoxyglucose. *Journal of Nuclear Medicine* 1988;29:1521-25.

Reisser C, Haberkorn U,Strauss LG. The relevance of positron emission tomography for the diagnosis and treatment of head and neck tumors.*Journal of Otolaryngology* 1993;22:231-8.

Zeitoni AG, Yamamoto L, Black M, Gjedde, A. Functional imaging of head and neck tumors using positron emission tomography. *Journal of Otolaryngology* 1994;23:77-80.

Appendix 4

Systematic Review: PET as a Diagnostic Test in Colorectal Cancer

Author: Karen Flynn, D.D.S., M.S., *Manager, MDRC Technology Assessment Program*

Appendix 4

Systematic Review:
PET as a Diagnostic Test in Colorectal Cancer

The final literature database searches for the systematic reviews were performed on September 10, 1996; the assessment represents peer-reviewed literature published and indexed as of that date.

This Appendix to the PET assessment presents the results of the systematic review of PET in colorectal cancer. A general rationale for the use of PET in oncology is supplied by Hawkins, et al. (1994) and Hoh, et al. (1994):

- many forms of cancer characteristically perturb tissue biochemical and physiological processes and PET imaging can be expected to detect the resulting abnormalities;
- reliance on tumor histology and anatomy limits the oncologist's tools for selecting optimal treatment;
- the ability to monitor metabolic responses to treatment could allow the early re-direction of therapy in patients who fail to respond to the first attempt at radiation or chemotherapy.

These and other authors (e.g., Price and Jones, 1995) report that PET studies in cancer are emerging as a major focus of the technology, both in basic research and in clinical investigations. Information gathered by the MDRC Technology Assessment Program from VA PET facilities corroborates that perception (see *Appendix 9: Experience With PET in VHA*).

Fluorine-18-fluorodeoxyglucose (FDG) is the most commonly employed radiopharmaceutical in PET cancer studies. Many neoplasms have high glycolytic rates, resulting in intracellularly trapped phosphorylated FDG that can be imaged with PET. Hawkins, et al. (1994), note that tumor-specific biochemical characteristics of glucose transport and phosphorylation may affect quantitative estimates of tumor glucose metabolism with FDG PET, and that investigations are under way to define these characteristics. However, these uncertainties may be of less concern with qualitative or semiquantitative FDG PET cancer studies because the primary intent of such studies is to detect and map tumor foci, not to rigorously quantify tumor glycolytic rates.

In some instances, PET imaging techniques have been modified to meet the needs of cancer diagnosis. Most PET systems allow axial fields of view (the length of the body encompassed by a series of cross sectional images) of approximately 10 cm. Cancer is frequently distributed beyond this field of view, and whole body image acquisition procedures have been developed (Hoh, et al., 1993). Since it is impractical to apply standard transmission scanning attenuation correction methods to these procedures, whole body PET imaging is primarily useful as a qualitative indicator of disease distribution.

Nieweg (1994) and Price and Jones (1995) define a number of potential applications for PET in oncology. These include:

- tumor detection (although PET images offer insufficient structural detail and should not be used to visualize anatomy; registration techniques to combine PET and anatomic imaging into a single image are under development to circumvent this limitation);
- staging (particularly using whole-body imaging methods) although there is a lower limit to the size of metastases that can be detected by PET;
- detection of local recurrence of disease, since anatomically-based imaging is often limited by the effects of treatment;
- prediction of tumor response to chemotherapy;
- treatment monitoring.

I. BACKGROUND

A. General sources

The material in this section, unless otherwise noted, is based on information in the National Cancer Institute's Physician Data Query (PDQ) system (retrieved in September, 1996).

B. Epidemiology

Colorectal cancer is recognized as a major source of morbidity and mortality, and a significant public health problem, for both men and women in the United States. Among cancer death rates, those for colorectal cancer are second only to lung cancer. These rates have fallen 29% for women, but only 7% for men, in the last 30 years.

Approximately 134,000 incident cases (10% of incident cases of all types of cancer) and 55,000 deaths (10% of all cancer related deaths) are estimated for the United States in 1996 (American Cancer Society, 1996). Within the Veterans Administration health care system, malignant neoplasms of the digestive organs and peritoneum (which include colorectal cancer) accounted for a total of 10,000 patients discharged (1% of all patients discharged within the system) with an average length of stay of 19.4 days (Annual Report of the Secretary of Veterans Affairs, 1994) during 1994.

Risk factors for colorectal cancer include age over 50 years, hereditary syndromes (familial polyposis of the colon and non-polyposis syndrome), and inflammatory bowel disease. There is some evidence that high fat and/or low fiber diets may also contribute to risk. The median age of colorectal cancer patients at diagnosis is 70 years, with less than 4% of cases occurring in patients younger than 50 years (Donald and Burhenne, 1993; US Preventive Services Task Force, 1996).

C. General description

Colorectal cancers are primarily of a single histologic type, adenocarcinoma. Metastases to the liver, abdominal cavity, and extra-abdominal areas at diagnosis are common, as is recurrent disease after surgical resection of the primary tumor. Prognosis is closely related to the depth of tumor penetration into the bowel wall and the presence of both regional lymph node involvement and distant metastases.

Most colon cancers are thought to develop from adenomas, but most adenomas do not progress to cancer. Potential approaches to the control of colon cancer include primary prevention by screening for and removing benign adenomatous polyps, and secondary prevention by screening for early cancer that might be removed with curative intent. Accordingly, a number of screening strategies are available, and continuing research into the efficacy of the various strategies is under way (US Preventive Services Task Force, 1996).

D. Staging, treatment, and survival

The prognostic variables noted in the paragraph above are incorporated into the Dukes staging system (Table 1). Unless gross evidence of metastatic disease is present, it is virtually impossible to determine accurately disease stage prior to surgical resection and pathologic analysis of operative specimens (Fengler and Pearl, 1994). Current surgical practices support more thorough intraoperative and pathologic staging than did previous practices; prognosis appears to be more precisely assessed not only through the presence or absence of metastatic nodes, but the number of involved nodes (i.e., 1 to 4 versus > 5).

The initial diagnosis of colorectal cancer is made on the basis of endoscopic or radiographic findings. Pre-operative evaluation includes: determination of carcinoembryonic antigen (CEA) levels [elevated titers (> 5.0 ng/mL) are associated with a high risk of metastatic disease and eventual tumor recurrence]; physical examination; a chest radiograph; and biochemical assessment of liver function. A colonoscopic evaluation of the entire large bowel may be performed to identify synchronous tumors or suspicious polyps. Computed tomography (CT) is widely used to determine the extent and location of lesions preoperatively. CT relies on anatomic changes to detect disease, and, like many other diagnostic tests, is most accurate when patients have advanced disease (Falk, et al, 1994). Other methods that may contribute to pre-operative staging include magnetic resonance imaging, transrectal ultrasound, laparoscopic exploration, and laparoscopic ultrasound (Falk, et al., 1994).

Table 1 summarizes the treatment options that are currently available to patients with colorectal cancer. As a general rule, surgical resection is the primary therapy. Various forms of adjuvant chemotherapy and radiation therapy are under investigation; some regimens, like combination chemotherapy with fluorouracil and levamisole after surgery for stage III disease, have been shown to be of benefit in certain settings. There is no standard treatment for advanced colorectal cancer; radiation therapy is under investigation. Chemotherapy appears to be of limited benefit in patients with advanced disease, and is associated with only a 15 to 20% chance of partial response or short-term palliation.

E. Follow-up after primary treatment

Most recurrences after surgical resection of colorectal cancer occur within the first 4 post-operative years; 80% of recurrences occur within 2 years (Bruinvels, et al., 1994). Therefore, many physicians observe patients carefully for up to 5 years by semiannual physical examinations, imaging studies, and yearly blood chemistry determinations, including CEA levels. A number of recent articles (Steele, 1993; Nelson, 1994; Kronborg, 1994) review the known costs, risks, and unquantified benefits of intensive follow-up strategies.

Two recently published randomized controlled trials found no survival benefit with intense follow-up versus no follow-up (with patients instructed to report for evaluation when they became symptomatic) (Ohlsson, et al., 1995) or with intense follow-up versus conventional follow-up (Makela, et al., 1995). An additional large-scale randomized trial to document the efficacy of a standard, postoperative monitoring program is in progress; the investigators hope to define any benefit associated with follow up strategies, and to define the subgroups of patients who would be most likely to benefit (Kronborg, 1988; Nelson, 1996).

Table 1 Modified Dukes classification of colorectal cancer, standard treatment options, and survival

Stage	Pathologic description	Colon		Rectum	
		Treatment	***5-year survival***	***Treatment***	***5-year survival***
A	Cancer limited to mucosa and submucosa	• wide surgical resection and anastomosis	> 90%	• wide surgical resection and anastomosis • selected patients may receive local resection with or without radiation plus chemotherapy	75-100%
B	Cancer extends into muscularis or serosa; uterus, parametria, ovaries, or prostate often involved	• wide surgical resection and anastomosis • patients should be considered for entry into carefully controlled trials evaluating the use of systemic or regional chemotherapy, radiotherapy, or biological therapy • adjuvant therapy not indicated unless patient participates in clinical trial • subgroups of patients at high risk for recurrence may be considered for adjuvant therapy	70-85%	• wide surgical resection and anastomosis when followed by chemotherapy and postoperative radiation • other surgical /adjuvant approaches within clinical trials • preoperative radiation with or without chemotherapy followed by surgery with attempt to preserve sphincter function with subsequent adjuvant chemotherapy, within a clinical trial	50-80%
C	Cancer involves regional lymph nodes	• wide surgical resection and anastomosis • patients who are not protocol candidates should receive postoperative chemotherapy (fluorouracil and levamisole) • eligible patients should be considered for entry into controlled trials comparing various postoperative chemotherapy regimens, radiation, or biological therapy, alone or in combination	30-60%	• wide surgical resection and anastomosis when followed by chemotherapy and postoperative radiation, preferably through participation in a clinical trial • other surgical /adjuvant approaches within clinical trials • preoperative radiation with or without chemotherapy followed by surgery with attempt to preserve sphincter function with subsequent adjuvant chemotherapy, within a clinical trial	30 - 60%
D	Distant metastases (liver, lung, etc.)	• surgical resection/anastomosis or bypass of obstructing lesions • surgical resection of isolated metastases to liver, lung, ovaries • palliative radio- or chemotherapy • clinical trials of new drugs and biologic therapy	< 10%	• surgical resection/anastomosis or bypass of obstructing lesions • surgical resection of isolated metastases to liver, lung, ovaries • palliative radio- or chemotherapy • clinical trials of new drugs and biologic therapy	<10%
Recurrent		• treatment depends on sites of recurrent disease demonstrated by physical examination and radiographic studies • isolated liver and lung metastases may be resected • palliative radio- or chemotherapy • patients candidates for phase I and II trials	5-year cure rate for resection of solitary or combination metastases > 20%; otherwise, prognosis poor	• treatment depends on sites of recurrent disease demonstrated by physical examination and radiographic studies • isolated lung or ovarian metastases may be resected • palliative radiotherapy • patients candidates for phase I and II trials of palliative chemotherapy	5-year cure rate for resection of up to 3 liver metastases > 20%, with some long-term cures; otherwise, prognosis poor

F. Potential roles for PET

The first report of potential roles for PET in colorectal cancer management was published by Yonekura, et al., from Brookhaven National Laboratory and the Memorial Sloan-Kettering Cancer Center, in 1982. These authors performed FDG PET studies on 3 patients with biopsy proven advanced liver metastases from colon cancer. All of the patients showed markedly increased accumulation of FDG in their liver tumors in images acquired late in the scan period (50 minutes after injection). FDG activity increased continuously in tumors following injection, while it decreased in normal liver tissue (tumor to normal liver ratios of 3.3 to 4.7). Yonekura, et al., concluded that FDG may be useful as an imaging agent for the detection and characterization of liver tumors.

Other authors (Beets, et al., 1994; Lai, et al., 1996; Vitola, et al., 1996) concur that estimating the resectability of liver metastases may be an area in which PET can have a significant clinical impact. In patients with apparently limited recurrent colorectal cancer (to the liver or lungs) 5-year survival rates of 20-30% can be obtained by resection with curative intent. Since these rates are only 20-30%, many of the patients who undergo surgery for resection must have unrecognized tumor foci. The morbidity (and costs) associated with surgery in patients who do not have genuinely resectable recurrent tumor could be avoided by improved methods of tumor detection.

Other potential roles for PET in colorectal cancer imaging have been identified. These include:

- pre-operative staging of disease (Falk, et al., 1994);
- postoperative monitoring of patients for recurrent disease (Strauss, et al., 1989; Ito, et al., 1992).

Other new nuclear medicine tests, such as monoclonal antibody imaging are also felt to be particularly useful in patients who have rising CEA levels during post-treatment monitoring, but no evidence of recurrence on conventional imaging studies such as CT or MRI, or in patients who are suspected to have an isolated, resectable recurrence and for whom surgery with curative intent is planned (Goldenberg, 1993; Peterson, et al., 1993; Tempero, et al., 1995).

II. RESULTS

Seventeen articles identified through MEDLINE and other database searches and from the bibliographies of initially retrieved articles were selected as meeting the screening criteria. After review, twelve (71%) met inclusion criteria: 5 met the definition of technical efficacy (listed in Section VII; full data abstraction tables for technical efficacy studies are on file with the MDRC Technology Assessment Program); and 5 met, to some extent, the evidence-based medicine criteria for diagnostic accuracy evaluations (Table 3). An additional 2 studies were classified at both the diagnostic accuracy and therapeutic efficacy levels (Table 4). A single study addressed only the effect of PET on treatment decisions and was also classified as a therapeutic efficacy study (Table 4).

The extent to which the potential applications of PET in colorectal cancer are supported by published evidence is indicated in Table 2, which details the methodologic quality of the evidence according to the potential role played by PET in colorectal cancer, and in Tables 3 and 4, which abstract data from studies classified at the diagnostic accuracy and therapeutic efficacy levels. PET appears to have very good face accuracy in distinguishing recurrent colorectal cancer from treatment artifacts such as scars, and in documenting hepatic or more distant metastases that might preclude surgery with curative intent. However, the methodologic limitations of the studies published to date should be taken into account when interpreting the accuracy data.

The studies in Tables 3 and 4 that address the detection of hepatic metastases may be associated with work-up bias, as PET and other imaging studies were used to direct biopsies to confirm the presence of malignancy in suspicious liver lesions. Sensitivity calculations (for both PET and alternative technologies) in such settings may be problematic and may overestimate accuracy, as the number of false negatives may not be accurately determined (Valk, 1996). While most authors made attempts to compensate for work-up bias (e.g., Lai, et al., reported that 20/34 patients had received intraoperative ultrasound to confirm the completeness of lesion identification and biopsy), there are limitation to the extent to which bias can be eliminated in this clinical setting (Stark, et al., 1987).

All of the studies in Table 3 were classified as case series, since patients were accrued as they presented for evaluation. Although the case series included some patients with benign (rather than malignant) lesions who could serve as internal controls, there was a lack of balance (sometimes of the order of 2 to 1) between numbers of cases and controls. Predictive values based on these case series would have substantial potential for bias. The small numbers of patients in the PET studies and the lack of documentation of disease severity among the cases would also argue for caution in interpreting and generalizing sensitivity and specificity data.

While the studies in Table 4 were classified at the therapeutic efficacy level, their results should also be interpreted and generalized with caution. These studies were retrospective case series that did not appear to have been specifically designed to document changes in treatment, methods for documenting such changes were not made explicit, and data tended not to be systematically analyzed or presented. The studies generally enrolled highly selected patients whose previous work-up was not clearly specified, nor was the size or composition of the referral base from which the patient sample was drawn. Information from PET studies resulted in more appropriate treatment for some patients. However, the published studies tended to give inadequate details about what happened to patients whose PET studies did not accurately reflect their disease status.

The potential role of PET in postoperative monitoring of patients for recurrent disease has not been addressed in the published literature, and would need to be evaluated in the context of the uncertain benefits of such monitoring (Kievet and Bruinvels, 1995; Makela, et al., 1995; Ohlsson, et al., 1995; Nelson, 1996). Finally, the MDRC Technology Assessment Program was unable to locate any studies that addressed the effect of incorporating PET into diagnostic strategies on patient outcomes or costs of care.

III. ALTERNATIVES TO PET AND DISCUSSION

Bruinvels, et al., (1994), note that the intensive follow-up strategies reported in the literature include a range of diagnostic technologies (physical examination, blood chemistry studies including CEA determinations, colonoscopy, barium enema, sigmoidoscopy, fecal occult blood, and liver ultrasound). Once a patient with a potential recurrent cancer has been identified by monitoring or by symptoms, other diagnostic tests, including PET and other nuclear medicine tests such as immunoscintigraphy (which uses a SPECT or gamma camera to image sites of localization for radio-labeled monoclonal antibodies directed against tumor cell antigens) have recently been identified as ways to estimate an individual patients suitability for potentially curative resection. Table 5 provides accuracy data for some of these tests for comparison with the information on PET (Table 3). All of these technologies should be evaluated in the context of the lack of evidence supporting intensive follow-up and the small percentage of patients for whom any second attempt at cure would be effective (Nelson, 1996).

In the context of the general uncertainty regarding postoperative screening and additional operations to treat recurrent disease, the published studies on PET provide only preliminary information. At the diagnostic accuracy level, PET studies have also failed to address the marginal benefits in accuracy obtained by PET relative to other technologies included in complex diagnostic strategies.

The MDRC Technology Assessment Program was unable to locate any studies that addressed the effect of incorporating PET into diagnostic strategies on patient outcomes or costs of care. Since a major rationale for using PET in evaluating patients with potentially resectable recurrent or metastatic disease is the avoidance of unnecessary surgery, impact of PET on survival and quality of life and on costs may be a particularly fertile area for future research.

IV. SUMMARY

Table 2 summarizes published findings on the diagnostic accuracy efficacy of PET and some of its alternatives for the diagnosis of colorectal cancer. All of the PET studies are retrospective case series, which provide Level V (the weakest) evidence regarding an association between the use of PET and improved patient outcomes.

Five diagnostic accuracy efficacy studies in Table 2 (Strauss, et al., 1989; Schlag, et al., 1989; Ito, et al., 1992; Vitola, et al., 1995; Lai, et al., 1995) met evidence-based medicine criteria for diagnostic test evaluations. These studies addressed the role of PET in differentiating recurrent cancer from scar and in diagnosing liver metastases. They were classified as “C” using other methodologic criteria, due to their relatively small numbers of cases, lack of equivalence between numbers of cases and internal controls, retrospective nature, incomplete descriptions of the “filters” through which patients passed to participate in the studies and severity of disease, and the presence of work-up bias. The lack of methodologic rigor in these studies, and its potential association with overestimation of PET’s diagnostic accuracy, would also apply to the alternative technologies to which PET was compared in the same studies. Diagnostic accuracy reports for alternative technologies that have been investigated using stronger study designs (e.g., the randomized study comparing CT and MRI in diagnosing liver metastases reported by Stark, et al., 1987) would be less subject to bias.

While data on other uses of PET are also included in Table 2, the MDRC Technology Assessment Program was unable to locate any published studies that met evidence-based medicine criteria for evaluations of diagnostic tests for the use of PET in these settings. Alternative technologies have been investigated for these settings with greater rigor; examples of more rigorous studies are included in Table 2.

Three studies (Table 4) have made attempts to address the role of PET in changing treatment decisions. These studies would be considered preliminary, due to their designs (retrospective case series that had not been specifically designed to document changes in treatment); methods for recording changes in treatment plans were not specified, and results data tended not to be systematically analyzed or presented. The studies generally enrolled highly selected patients whose previous work-up was not clearly specified, nor was the size or composition of the referral base from which the patient sample was drawn. Information from PET studies resulted in more appropriate treatment for some patients. However, the published studies tended to give inadequate details about what happened to patients whose PET studies did not accurately reflect their disease status.

Table 2 Summary of the literature: Diagnostic accuracy efficacy of PET and alternatives in colorectal cancer

Notes *The PET studies in this table were retrospectively analyzed case series; internal controls (cases with benign, rather than malignant, conditions) allowed the calculation of specificity as well as sensitivity.*

Some studies analyzed results separately according to the clinical role of PET for subsets of patients; these studies appear in the table more than once, and may have received different methodologic quality grades for each subset analysis.

Role	Study	N	Operating characteristics*				Evidence-based medicine criteria**			Methodologic quality grade***
			PET	*CT*	*MRI*	*Other*	*controls*	*standard*	*blinding*	
Detecting or staging primary or recurrent disease	Falk, et al., 1994	16 patients: 15 malignant lesions; 3 benign lesions	Se = 87% Sp = 67%	Se = 47% Sp = 100%			+ (internal)	+	partial	D
	Nattinger, et al., 1991 (ACP review)					colonoscopy Se = 94% Sp = 100%	(review)	(review)	(review)	(review)
	Hernandez-Socorro, et al., 1995	40 cases 64 controls				colonoscopy Se = 94% Sp = 100% hydrocolonic ultrasound Se = 97% Sp = 97%	+	+	+	B
Diagnosing recurrent tumor vs scar	Strauss, et al., 1989	29 patients: 21 malignant lesions; 8 scar	Se = 95% Sp = 100%				+ (internal)	+	+ (quantitative analysis)	C
	Schlag, et al., 1989	18 patients: 11 malignant lesions; 6 scar	Se = 92% Sp = 100%			immunoscintigraphy Se = 40% Sp = 50%	+ (internal)	+	+ (quantitative analysis)	C
	Ito, et al., 1992	15 patients: 11 malignant lesions; 4 scar	Se = 100% Sp = 100%		Se = 91% Sp = 100%		+ (internal)	+	+ (quantitative analysis)	C
	Schiepers, et al., 1994	6 patients: 5 malignant lesions; 1 scar	Se = 100% Sp = 100%				+ (internal)	+	-	D
Diagnosing recurrent tumor vs scar	Hawes, et al., 1993	85 with disease 408 without disease (review with weighted average of results from 7 studies)				endoscopic ultrasound Se = 99% Sp = 88%	(review)	(review)	(review)	(review)

Role	Study	N	Operating characteristics*				Evidence-based medicine criteria**			Methodologic quality grade***
			PET	*CT*	*MRI*	*Other*	*controls*	*standard*	*blinding*	
Diagnosing liver metastases	Schiepers, et al., 1994	80 studies: 34 malignant lesions; 46 benign lesions	Se = 94% Sp = 100%			CT and/or ultrasound Se = 85% Sp = 98%	+	+	-	C
	Vitola, et al., 1996	55 sites: 39 malignant; 16 benign 24 patients: 19 malignant disease; 5 benign	Se = 90% Sp = 100% Se = 95% Sp = 100%	Se = 86% Sp = 58%		CT portography Se = 97% Sp = 9% Se = 100% Sp = 33%	+ (internal)	+	+ (semiquantitative analysis)	C
	Lai, et al., 1996	34 patients: 27 malignant disease; 7 benign or no disease	Se = 93% Sp = 57%	Se = 100% Sp = 14%	Se = 100% Sp = 80%		+ (internal)	+	+	C
	Stark, et al., 1987	57 cases; 72 controls: 21 benign liver disease; 51 with normal livers		Se = 80% Sp = 94%	Se = 82% Sp = 99%		+	+	+	B
	Panzer, et al., 1991 (ACP review)	review		Se = 90% Sp = 90% LR + = 8 LR - = 0.11		ultrasound, adequate studies Se = 80% Sp = 90% LR + = 9 LR - = 0.22	(review)	(review)	(review)	(review)
Diagnosing liver metastases	Rafaelsen, et al., 1995	295 patients: 64 with liver metastases 231 without liver metastases				liver enzymes Se = 9-47% (ALT 9%, AKP 31%, LDH 47%) Sp = 92-98% preop US Se = 70% Sp = 94% surgical exploration Se = 84% Sp = 97% intraop US Se = 97% Sp = 98%	+ (internal)	+	+	B

Abbreviations

CT, computed tomography
MRI, magnetic resonance imaging
neg, negative for disease
pos, positive for disease
LR, likelihood ratio
AKP, alkaline phosphatase
PPV, positive predictive value
NPV, negative predictive value
US/FNA, ultrasound/fine needle aspiration
ACP, American College of Physicians
ALT, alanine amino transferase
LDH, lactate dehydrogenase

*operating characteristics defined in Appendix 2: Assessing Diagnostic Technologies, page 5-7
**Appendix 2, page 8
*** Appendix 2, page 9

V. SUGGESTIONS FOR FUTURE RESEARCH

Existing research supplies a preliminary indication of PET's potential benefit in the diagnosis and management of colorectal cancer patients. The studies that have been published had design limitations and enrolled small numbers of patients. These studies have made preliminary attempts to define the operating characteristics of PET as a diagnostic test, particularly in the setting of follow-up strategies to detect recurrent disease, and to document changes in treatment based on PET results.

1) Contributions from larger patient populations and stronger study designs are needed to refine the characteristics of PET as a diagnostic test in colorectal cancer, and to establish a base for further research.

2) A PET registry could provide a range of data on demographic and clinical characteristics of patients on whom PET studies are performed, and on their clinical outcomes in a variety of settings

3) The use of PET to avoid unnecessary surgery by detecting unresectable recurrent disease in patients who are scheduled for surgery based on other imaging and blood chemistry studies should be documented more systematically and in larger patient samples.

4) If the ongoing randomized clinical trial (Kronborg, 1988) indicates that postoperative follow up in colorectal cancer patients reduces mortality, the marginal gains attributable to PET when it is incorporated into a multi-test follow-up strategy should be quantified.

Table 3 ***Data abstraction table:***
Diagnostic accuracy efficacy of FDG PET and alternative technologies in colorectal cancer

Notes: Studies in this table were designed to evaluate the use of PET in diagnosing primary or recurrent colorectal cancer (Falk, et al.) or in distinguishing recurrent cancer from treatment artifact (e.g. scar). All of the studies met , to some degree, the evidence-based criteria for diagnostic test evaluations. All would be classified as case series, since they accrued subjects as patients were referred for evaluation of suspected cancer. However, the cases included patients with benign (as opposed to malignant) masses, and these patients can serve as internal controls. Other methodologic limitations are discussed in the text and noted for each study.

Unless otherwise noted, the studies in this table compared PET and CT or MRI to the "gold standard" of histopathology of surgical specimens, which is the reference test for the operating characteristics reported in the "Results/Comments" column.

Study	***Patients/Methods***	***Results/Comments***
Strauss, et al., 1989 *German Cancer Research Center, Heidelberg*	**Purpose** to differentiate recurrent colorectal cancer from scar tissue **Cases** 29 patients with suspected local recurrence of colorectal cancer (21 malignant lesions/8 nonmalignant) **Methods** • all subjects examined with both FDG and ^{15}O-water PET • due to limited resolution of PET, only patients with lesions exceeding 1.5 cm diameter on CT were included • final diagnosis based on biopsy and FU • PET examinations were performed after a final diagnosis had been obtained by means of biopsy and/or consistency of sequential CT findings • PET images analyzed quantitatively by means of ROIs and DAR with gluteal muscles serving as normal reference tissue **Study design limitations** • numbers of cases and internal controls not equivalent (high prevalence of malignancy) • cut off values for normal/abnormal tests not specified	**FDG uptake** • rapid FDG uptake by tumor, followed by slight decrease in DAR for up to 40 minutes after administration; FDG concentration in tumor at 1 hour post-injection was > 2x that in scar • FDG uptake at 60 minutes was low in nonmalignant lesions • tumor/scar FDG uptake ratio was best at 60 minutes after injection • DAR values for normal tissues were constant for 50 minutes, beginning at 10 minutes post-injection **DAR values** • recurrent tumor = 1.14 - 4.17 • scar = 0.56 - 1.15 **Lesion/soft tissue FDG uptake ratios** • recurrent tumor median ratio = 2.08 • scar median ratio = 0.96 **Recurrent tumor vs scar based on FDG DAR and lesion/soft tissue ratio values** PET: *Se = 95%, *Sp = 100%

Study	*Patients/Methods*	*Results/Comments*
Schlag, et al., 1989 *German Cancer Research Center, Heidelberg*	**Purpose** investigate the feasibility and utility of PET and immunoscintigraphy in distinguishing recurrent rectal cancer from scar tissue **Cases** 18 patients with clinically suspected recurrence of rectal cancer • lesions ≥ 1.5 cm on CT • 11 malignant/6 nonmalignant **Methods** • all subjects received PET • 14/18 had elevated CEA levels and received immunoscintigraphy with radioactive-labelled CEA/Ca 19-9 • PET images analyzed quantitatively using ROIs and time-activity curves; DAR calculated with gluteal muscles as normal soft tissue reference **Study design limitation** numbers of cases and internal controls not equivalent (high prevalence of malignancy)	**PET** • mean FDG tumor to normal soft tissue ratio = 2.7 • differentiation best at 57.5 minutes after FDG injection • Se = 92%; Sp = 100% for tumor vs scar **Immunoscintigraphy with CEA/Ca 19-9** • Se = 40%; Sp = 50% for tumor vs scar • less accurate than other reports; attributed to binding and specificity (for colorectal cancer) of the antibodies and using histopathology as gold standard **Authors' comment** PET could improve selection of patients for invasive procedures
Ito, et al., 1992 *Nagoya University School of Medicine, Japan*	**Purpose** • to compare the value of PET and MRI in differentiating recurrent rectal carcinomas from scars • to investigate the role of PET in determining response to therapy **Cases** 15 patients with suspected (abnormal CT) local recurrence of rectal cancer (11 malignant/4 nonmalignant) **Methods** • final diagnosis obtained by surgery in 2 patients, biopsy in 7 patients, CT of bone destruction in 2 patients, and sequential CT in 2 patients • procedures to minimize FDG activity in urinary bladder used • 7 patients received sequential PET studies after treatment • MRI obtained and superimposed on all PET images • FDG uptake quantified by DAR in ROIs **Study design limitations** • numbers of cases and internal controls not equivalent (high prevalence of malignancy) • cut off values for normal/abnormal tests not specified	**PET** • mean DAR values significantly different for recurrent tumor vs scar ($p < .01$) **MRI** • lesion/muscle signal intensity ratios for recurrent tumors and scars were significantly different ($p < .01$) • DAR and signal intensity ratios correlated ($r = .603$; $p < .05$) **Distinguishing recurrent rectal tumor from mature scar** • PET: *Se = 100%; *Sp = 100% • MRI: *Se = 91%; *Sp = 100% (1 patient who was positive for disease at surgery misclassified by both MRI and biopsy) **Treatment monitoring** • 6/7 patients had decrease in DAR by completion of radiation therapy; in 4 of the 6 tumor did not decrease in size • 1/7 patients did not respond to treatment, and both DAR and tumor size increased during treatment • no longer term follow up results noted **Authors' conclusion** PET and MRI are complementary, particularly where the results of MRI are atypical

Study	Patients/Methods	Results/Comments
Falk, et al., 1994 *Creighton University School of Medicine, Omaha, and West Virginia University School of Medicine*	**Purpose** to determine the sensitivity, specificity, and predictive accuracy of PET and CT preoperatively in patients with colorectal cancer **Cases** 16 patients with suspected or biopsy-proven primary or recurrent colorectal cancer • 15 malignant lesions confirmed at biopsy (12 sites in colon and rectum, 2 liver metastases, 1 mesenteric metastasis) • 3 nonmalignant lesions **Methods** • all subjects received PET and CT after at least 4 hours of fasting • qualitative analyses performed • image reviewers blinded to results of other imaging studies; blinding to other test results not noted **Study design limitation** numbers of cases and internal controls not equivalent (high prevalence of malignancy)	**PET** • detected 12/12 malignant lesions in colon and rectum and 1/2 liver sites • 1 false positive scan attributed to inflammation • 2 false negative scans; 1/2 attributed to lack of clear demarcation between liver and right colon • Se = 87%; Sp = 67%; accuracy = 83% **CT** • lesions missed generally quite large (≥ 25 mm) • Se = 47%; Sp = 100%; accuracy = 56% **Other findings/authors' conclusions** • no PET or CT related complications observed • PET and CT are complementary; PET may be especially useful if CT findings equivocal • PET costs approximately twice those of CT but may be justified if unnecessary surgery prevented or unexpected early lesion detected
Schiepers, et al., 1995 *University Hospital Gasthuisberg, Belgium*	**Purpose** to evaluate contribution of whole-body PET to detecting and localizing local recurrence and metastatic disease, compared to CT-pelvis and CT/ultrasound-liver **Cases** 74 consecutive patients presenting for evaluation of suspected recurrent disease at median 1 year post surgery (45 with recurrent disease, 29 with benign conditions) • final diagnosis by biopsy in 63% • final diagnosis by FU ≥ 14 months in 37% **Methods** • work up after recurrence suspected included CEA/CA 19.9, CT-pelvis, ultrasound or CT-liver, chest x-ray, colonoscopy • 83 PET studies in 74 patients • PET interpreted qualitatively • operating characteristics calculated by site (not patient) **Study design/reporting limitations** • blinding not noted • validation of tumor in liver and distant sites dependent on imaging results	**All recurrences: PET vs CT** (74 studies: 45 malignant, 29 benign) • PET: Se = 93%; Sp = 97%, accuracy = 95% • CT: Se = 60%; Sp = 72%, accuracy = 65% • 1% of PET and 15% of CT studies equivocal; equivocal studies counted as false positives or false negatives in accuracy calculations **Local fibrosis vs recurrence: PET vs CT** (6 patients: 5 recurrent, 1 scar) PET: Se = 100%; Sp = 100% **Liver involvement: PET vs CT and/or ultrasound** (80 studies: 34 malignant, 46 benign) • PET: Se = 94%; Sp = 100%, accuracy = 97% • CT/ultrasound: Se = 85%, Sp = 98%, accuracy = 92% **Distant extrahepatic disease: PET** • PET detailed 25 unexpected lesion locations in 20 patients; 14 (56%) lesions confirmed with biopsy or other imaging • absence of disease in all false positives (all in thorax) confirmed with long term FU • 1 false negative confirmed at surgery

Abbreviations:
DAR, differential absorption ratio
FU, follow up
Se, sensitivity
Sp, specificity
PPV, positive predictive value
NPV, negative predictive value
ROI, region of interest
SUV, standardized uptake value
T/B, target-to-background ratios
CEA, carcinoembryonic antigen

* indicates calculated by MDRC TA Program from data supplied in published article

Table 4 ***Data abstraction table:***
Therapeutic efficacy of FDG PET in colorectal cancer

Notes: Studies in this table provide both diagnostic accuracy and some therapeutic efficacy results. The degree to which the studies meet evidence-based criteria for diagnostic test evaluations and other methodologic criteria is variable (see Table 2).

All of the studies listed here are retrospectively analyzed case series. However, the cases included patients with benign (as opposed to malignant) masses, and these patients served as internal controls.

Sensitivity, specificity, and accuracy data presented here should be interpreted with caution, since validation of tumor in liver and distant sites was dependent on imaging results.

Therapeutic efficacy results are based on highly selected patients who had received a variable number and type of other diagnostic tests prior to PET (PET was complementary to, rather than an alternative to, other diagnostic tests); neither the number of patients who entered the diagnostic process at each institution nor the size of the referral base for the institution are specified, making generalization of the results presented here problematic. "True" PET results made positive contributions to treatment in these studies. False negative and false positive PET results may also have had negative impacts on some patients.

Study	***Patients/Methods***	***Results/Comments***
Vitola, et al., 1996 *Vanderbilt University Medical Center*	**Purpose** to compare whole-body PET, CT, and CT portography in detecting hepatic metastases **Cases** 24 patients presenting for evaluation of suspected (increasing CEA levels or abnormal CT) colorectal ca recurrence > 1 year after surgery (19 with recurrent ca, 5 with benign lesions) • 55 intrahepatic sites (39 malignant, 16 benign) • 5 extrahepatic sites (4 malignant, 1 benign) • diagnosis confirmed by histopathology in 19 patients, 1 year FU in 5 patients **Methods** • 17 patients had CT, 18 had CT portogram, 11 had both • all patients had PET, which was analyzed semiquantitatively using ROIs, T/B ratios and SUV • if PET showed extrahepatic lesion, additional CT in that area performed • selection of liver biopsy sites based on CT portography, which detected the greatest number of lesions	**Diagnostic accuracy efficacy** **PET** cut points: SUV = 3.5; T/B = 2 • by site: Se = 90%; Sp = 100% • by patient: *Se = 95%; *Sp = 100% **CT** • by site: Se = 86%; Sp = 58% • by patient: insufficient information for calculations **CT portography** • by site: Se = 97%; Sp = 9% • by patient: *Se = 100%; *Sp = 33%
	Study design/reporting limitations • extrahepatic lesion analysis not presented here due to work up bias in selecting patients for CT or other diagnosis on the basis of PET • validation of tumor in liver and distant sites dependent on imaging results • operating characteristics calculated completely only for site (not patient) as unit of analysis; clinical decisions made on patient basis • high prevalence of recurrent ca • treatment impact of false negative PET studies not discussed	**Therapeutic efficacy:** **alterations to treatment based on PET diagnosis** • 4/24 (17%) patients with hepatic metastases: - negative PET led to avoiding unnecessary laparotomy in 2 patients - positive PET led to partial hepatectomy in 2 patients who would not otherwise have received the procedure (other studies were false negative) • 2/5 (40%) patients with extrahepatic metastases

Study	Patients/Methods	Results/Comments
Lai, et al., 1996 *Royal Prince Alfred Hospital, Australia*	**Purpose** to compare PET with abdominal CT, chest CT, chest x-ray in identifying operable colorectal ca metastases to liver **Cases** 34 consecutive patients referred for evaluation of suspected metastases to liver • 27 malignant, 7 benign or no disease • diagnosis confirmed by histopathology in surgical specimens, percutaneous biopsy, serial CT (median FU, 18 months), intraoperative ultrasound of liver **Methods** • all patients had staging by abdominal CT, and plain film chest x-ray (15) or CT (19) • patients whose metastases were considered operable received MRI (24) or CT angiography (3) • conventional imaging studies interpreted by 2 senior radiologists blinded to PET results • PET performed after conventional imaging • PET interpreted qualitatively by single observer blinded to conventional imaging results **Study design/reporting limitations** • data insufficient to reproduce comparison of CT vs PET for extrahepatic metastases • number of cases and internal controls not equivalent: high prevalence of disease and inability to calculate predictive values • data insufficient for analysis of subset of patients who received intraoperative ultrasound for confirmation of imaging results separately from other patients • work up bias in selecting patients for MRI • validation of tumor in liver and distant sites dependent on imaging results • methods of evaluating changes from pre- to -post-PET treatment plans not specified	**Diagnostic accuracy** **Detection of hepatic metastases** • PET: *Se = 93%; *Sp = 57%; - one case of multiple metastases not detected on MRI was identified by PET - false positives in liver cysts • CT: *Se = 100%; *Sp = 14% • MRI: *Se = 100%; *Sp = 80% **Authors' comments** • PET is more sensitive than CT in detecting extrahepatic metastases (below), and has become the initial examination of choice for patients with presumed recurrent colorectal metastases to the liver at this institution • cost of PET is justified if unnecessary tests, hospital admissions, and surgery are avoided • semiquantitative (vs qualitative) whole body PET analysis may improve accuracy **Therapeutic efficacy** **Detection of extrahepatic metastases** • PET identified previously unsuspected lesions (missed by conventional imaging) in 11 patients • clinical management influenced by PET in 10 patients (29% of total evaluated) • 1 false positive PET (retroperitoneal nodes) • 1 false negative PET (para-aortic nodal metastases apparent at repeat PET in 1 year) • 3 equivocal PET findings (poorly localized FDG uptake in area of left hepatic lobe)
Beets, et al., 1994 *University Clinic Gasthuisberg, Belgium*	**Purpose** to evaluate clinical impact of whole-body PET in detecting and localizing recurrent colorectal ca **Cases** 35 patients with suspected recurrent disease who had received the following diagnostic battery below, up to and including PET **Methods** after surgery: • 6 monthly FU with clinical exam, serum CEA determinations, US liver, chest x-ray; colonoscopy at 1 year and then every 3 years • if recurrence suspected or identified, patients then had CT of pelvis, US and/or CT of liver, chest x-ray, colonoscopy, endorectal ultrasound • if recurrence still suspected or identified, patients then had pelvic MRI and CT of thorax • if results still equivocal, patients had PET **Study design/reporting limitations** • referral base of institution and number of patients screened or entered into diagnostic work up not specified • reported as 35 individual case reports; data not systematically analyzed • incomplete details on patients who had false positive or false negative PET results: treatment received/not received?	**Therapeutic efficacy:** **16 patients considered before PET to have resectable liver (15 or lung (1) metastases** • 9 patients: no additional information supplied by PET • 3 patients with equivocal CT (local anastomotic recurrence not ruled out): PET negative and patients had surgery to resect metastases • 4 patients: PET positive for advanced liver involvement and resection not attempted **8 patients considered before PET to have resectable local recurrence; PET used to detect additional sites which would rule out surgery** • PET correctly identified all local recurrences • 1 patient: surgery avoided due to unexpected pulmonary metastases on PET • 5 patients: PET had no therapeutic impact • 2 patients: false negative PET studies for metastases (metastases detected at surgery) **8 patients with presacral mass equivocal on CT** • 5 patients: diagnosis with PET accurate (1 true negative, 4 true positives) • 1 patient: false negative PET (malignancy discovered later) • 2 patients: false positive PET **3 patients with increasing CEA but no other evidence of recurrence** • 2 patients: PET correctly identified pelvic recurrence and patients had treatment • 1 patient: false negative PET (biopsy confirmed recurrence later)

Abbreviations: DAR, differential absorption ratio; FU, follow up; Se, sensitivity; Sp, specificity; PPV, positive predictive value; NPV, negative predictive value; ROI, region of interest; SUV, standardized uptake value; T/B, target-to-background ratios; CEA, carcinoembryonic antigen

* indicates calculated by MDRC TA Program from data supplied in published article

Table 5 ***Summary of the literature***
Diagnostic accuracy of alternative technologies to PET in colorectal cancer

Note: This table includes information from review performed for the American College of Physicians (ACP), as well as studies reporting primary data. The ACP reviews provide an overview of the accuracy of commonly used diagnostic tests for colorectal cancer. Many of these studies also provide models of more methodologically rigorous study designs than those that have been used in evaluating PET (see Table 2).

Study/design	***Patients/Methods***	***Results/Comments***
Diagnosing and/or staging primary colorectal cancer		
Nattinger, 1991 *ACP*	review	**Flexible sigmoidoscopy** • 60 cm scope: Se for cancer = 30 - 50%; Se for adenoma = 50 - 60%; Sp (any neoplasia) = 97% • 35 cm scope: Se for cancer = 40 - 50%; Se for adenoma = 40 - 50%; Sp (any neoplasia) = 100% **Air contrast barium enema** Se for cancer = 82 - 92%; Se for adenoma = 50%; Sp (any neoplasia) = 95% **Colonoscopy** Se for cancer = 94%; Se for adenoma = 94%; Sp (any neoplasia) = 100%
Hernandez-Socorro, et al., 1995 *Hospital del Pino, Canary Islands, Spain* diagnostic accuracy efficacy case series (Level V evidence)	**Purpose** to determine sensitivity and specificity of hydrocolonic sonography (transabdominal sonography after retrograde instillation of water into the colon) in detecting and staging colon cancer **Cases** 104 subjects referred for evaluation of colorectal disease • 40 malignant tumors (35 primary colorectal, 2 recurrent, 1 metachronous, 1 synchronous, and 1 metastatic uterine cancer) • 64 nonmalignant **Methods** • al subjects received conventional abdominal or endorectal sonography and hydrocolonic sonography prior to colonoscopy and single- or double-contrast barium enema • disease status verified by histology of resected surgical specimens • image interpreters blinded to colonoscopy, barium enema and histology results	**Hydrocolonic sonography** (40 cases, 64 controls) • primary or recurrent colon cancer: Se = 97.5%; Sp = 98.4%; PPV = 97.5%; NPV = 98.4% • tumor staging: 100% of T1, T2, T4 tumors and 96% of T3 tumors correctly classified • presence or absence of peritumor metastatic lymph nodes ≥ 4 mm: Se = 50%; Sp = 100% **Colonoscopy** (36 cases, 45 controls) Se = 94.4%; Sp = 100%; PPV = 100%; NPV = 95.7% **Barium enema** • single (32 cases, 64 controls): Se = 93.7%; Sp = 98.4%; PPV = 96.7%; NPV = 96.8% • double (11 cases, 19 controls): Se = 100%; Sp = 94.7%; PPV - 91.6%; NPV = 100% **Conventional ultrasonography** (40 cases, 64 controls) Se = 40%; Sp = 98.4%; PPV = 94.1%; NPV = 72.4
Evaluating suspected recurrent disease		
Hawes, 1993 *Indiana University Hospital*	**Purpose** review of endoscopic ultrasound accuracy in distinguishing recurrent rectal cancer from scar **Methods** summary Se and Sp from 7 studies (85 cases/408 controls) using weighted averages	**Endoscopic ultrasound** Se = 99% Sp = 88%

Study/design	*Patients/Methods*	*Results/Comments*
Gasparini, et al., 1994 diagnostic accuracy efficacy case series (Level V evidence)	**Purpose** to compare immunoscintigraphy with anti-CEA monoclonal antibody to CT, ultrasonography, and MRI in patients with suspected local recurrence of colorectal cancer **Cases** 59 patients referred for evaluation of suspected recurrence (≥ 2 consecutive increases in CEA levels) • 45 with recurrence by histology or endoscopy • 14 with benign lesions by follow up or histology **Methods** • whole body and multiple regional spot gamma camera scintigraphic images obtained at 4, 24, 48, and 72 hours after antibody injection • SPECT images obtained at 48 - 72 hours • only pelvic sites analyzed **Study design limitations** • not all subjects had histologic confirmation of disease status • blinding of image readers not noted	**Immunoscintigraphy** Se = 89%; Sp = 78%; accuracy = 86% **MRI** Se = 93%; Sp = 67%; accuracy = 86% **CT** Se = 69%; Sp = 67%; accuracy = 68% **Ultrasound** Se = 41%; Sp = 79%; accuracy = 56%
Corman, et al., 1994 *Sansum Clinic, CA; University of Louisville; University of Chicago; Buffalo VAMC; University of Missouri; Cytogen Corp.* diagnostic accuracy and therapeutic efficacy case series (Level V evidence)	**Purpose** to assess diagnostic accuracy, and contribution to diagnostic thinking and subsequent treatment decisions of FDA-approved immunoscintigraphy agent (*Oncoscint*) **Cases** 103 patients (84 with confirmation of diagnosis by histology or other tests, 103 supplied data on contribution to diagnostic understanding or to subsequent therapeutic decisions) • 46 with rising CEA levels and otherwise negative evaluations • 29 with known recurrence, presumed resectable • 28 with equivocal results after other diagnostic tests **Methods** • gamma camera scintigraphic images of pelvis, abdomen, thorax and other sites obtained at 48 to 72 hours after antibody injection • images analyzed by nuclear medicine physician at each of 10 participating centers before confirmation of diagnosis by other means • clinicians provided description of pre-test management plan and assessed changes in plan due to immunoscintigraphy after surgery or at completion of diagnostic evaluation **Study design limitations** • not all subjects had histologic confirmation of disease status • Se and Sp calculations difficult to reconstruct	**Accuracy** • based on 84 patients with diagnosis confirmed by surgery, other tests, or follow up • scans indeterminate in 19/84 patients (23%); excluded from authors' accuracy calculations • Se = 73%; true negative rate = 100% (no positive scans in patients without other evidence of malignance; indeterminate tests not used in calculations) • if indeterminate scans considered false positive, Sp = 64% (MDRC TA Program calculation) **Effect on treatment** • beneficial in 44% of cases: 17 treatment plans altered due to detection of occult disease (disease not resectable and surgery canceled, surgery changed to radiation or chemotherapy, or surgical plan changed) • detrimental in 2% (test results led to unnecessary surgery in attempt to identify suggested recurrence)

Study/design	Patients/Methods	Results/Comments
Detecting hepatic metastases from colorectal cancer		
Stark, et al., 1987 *Massachusetts General Hospital, Boston* diagnostic accuracy efficacy randomized controlled study (Level II evidence)	**Purpose** to determine the accuracy of MRI (individual pulse sequences and combined sequences) relative to CT (contrast enhanced) in the diagnosis of liver metastases **Cases** 57 patients with biopsy-proven primary cancer (24 colon, variety of others) and liver metastases (proven by biopsy in 23 and by FU in 34) • all had both CT and MRI **Controls** • 27 patients with benign liver disease (17 had MRI only) • 51 subjects with normal livers (11 had CT only, 17 MRI only, 23 both) **Methods** • 438 MRI and 97 CT studies placed in individual folders, no patient identifiers on films or folders, folders randomized by investigator who did interpret studies • 3 blinded investigators independently interpreted studies • MRI studies reaggregated for 124 patients, rerandomized, and reinterpreted by 3 investigators independently • images analyzed by patient and by lesion and recorded on score sheets • final diagnosis on abnormality and number of lesions by consensus among all investigators with all information available • ROC analysis of test performance using data from patients with metastases or normal livers (not those with benign conditions) • differences between results of several MRI techniques and CT tested statistically **Study design/reporting limitations** • false negatives and false positives analyzed for both tests, but interobserver variability described but not quantified by kappa statistic • some subgroup analyses based on small numbers of cases and/or controls	**ROC analysis** • average area under curve for MRI larger than that for CT • all 3 interpreters operated on nearly same curve for both MRI and CT, but dispersion greater for CT (wider range of performance) • for all readers, optimal performance achieved when “probably abnormal” studies scored as negative **Performance characteristics** using cut points from upper left corner of ROC curves: • Se for abnormalities: MRI = 82%; CT = 80% • Sp: MRI = 99%; CT = 94% **Detection of individual metastatic lesions** (279 lesions in 39 patients who had both MRI and CT) • Se: MRI = 64%; CT = 51% **Detection of benign liver disease** • MRI: Se (hemangiomas) = 80%; Se (cysts) = 76% • CT: Se (hemangiomas) = 74%; Se (cysts) = 64% **Detection of extrahepatic lesions** • pancreatic masses - MRI: Se = 27%; Sp = 99% - CT: Se = 79%; Sp = 96% • renal cyst or mass - MRI: Se = 27%; Sp = 99% - CT: Se = 67%; Sp = 99% • adrenal mass - MRI: Se = 22%; Sp = 100% - CT: Se = 56%; Sp = 99% • focal splenic lesions - MRI: Se = 0%; Sp = 100% - CT: Se = 33%; Sp = 99% • ascites - MRI: Se = 83%; Sp = 98% - CT: Se = 50%; Sp = 98%

Study/design	Patients/Methods	Results/Comments
Rafaelsen, et al., 1995 *Odense University Hospital, Denmark* diagnostic accuracy efficacy prospective cohort study (Level III evidence)	**Purpose** to compare diagnostic accuracy of liver enzyme determinations, preoperative ultrasound, surgical examination, and intraoperative ultrasound for detection of liver metastases from colorectal cancer **Cases** 295 consecutive patients (1989 to 1992) admitted for elective surgery for colorectal cancer **Methods** • all patients received preoperative US, liver enzyme measurement, inspection of liver during surgery (findings recorded before intraoperative US), intraoperative US • surgeon, preoperative ultrasonologist, intraoperative ultrasonologist all unaware of each others' findings • disease status confirmed by combination of all tests (after recording of blinded findings), histopathology, and 3 mo postoperative FU (conventional US of liver, biopsy of liver if metastases suspected) • test characteristics calculated by patient and by lesion • differences between test results tested statistically	**Surgical procedures and staging results** • 216 curative operations, 79 palliative operations • Dukes stages A (35), B (148), C (33), D (64) **Detection of liver metastases (295 patients)** • liver enzymes: Se = 9 - 47%, Sp = 92 - 98%, depending on specific assay • preoperative US: Se = 70%; Sp = 94% • surgical exploration: Se = 84%; Sp = 97% • intraoperative US: Se = 97%; Sp = 98% significant differences: intraoperative US vs all other tests **Detection of unresectable metastases (46 patients with bilobar metastases, 35 > 3 lesions)** • intraoperative US: 91% of patients with bilobar metastases; 89% of patients with > 3 metastases • surgical exploration: 72% of bilobar metastases; 66% of patients with > 3 **Analysis by lesions (204 metastatic lesions in 64 patients)** • preoperative US: Se = 64%; Sp = 92% • surgical exploration: Se = 72%; Sp = 96% • intraoperative US; Se = 94%; Sp = 98%
Panzer, 1991 *ACP*	review	**Ultrasound, adequate studies** Se = 80%; Sp = 90%; LR positive = 8; LR negative = 0.22 **Computed tomography** Se = 90%; Sp = 90%; LR positive = 9; LR negative = 0.11

Abbreviations:
ACP, American College of Physicians
CEA, carcinoembryonic antigen
FU, follow-up
Se, sensitivity
Sp, specificity
PPV, positive predictive value
NPV, negative predictive value
LR, likelihood ratio
US, ultrasound

VI. REFERENCES: General background and diagnostic accuracy efficacy studies

Collier BD, Abdel-Nabi H, Doerr RJ, Harwood SJ, Olsen J, Kaplan EH, et al. Immunoscintigraphy performed with In-111-labeled CYT-103 in the management of colorectal cancer: comparison with CT. *Radiology* 1992;185:179-86.

Charnsagavej C. New imaging modalities for follow-up of colorectal carcinoma. *Cancer* 1993; 71:4236-40.

Corman ML, Galandiuk S, Block GE, Prager ED, Weiner GJ, Kahn D, et al. Immunoscintigraphy with ^{111}In-satumomab pendetide in patients with colorectal adenocarcinoma: performance and impact on clinical management. *Diseases of the Colon and Rectum* 1994;37:129-37.

Donald JJ, Burhenne HJ. Colorectal cancer: Can we lower the death rate in the 1990s? *Canadian Family Physician* 1993;39:107-14.

Falk PM, Gupta NC, Thorson AG, Frick MP, Bowman BM, Christensen MA, et al. Positron emission tomography for preoperative staging of colorectal carcinoma. *Diseases of the Colon and Rectum* 1994;37:153-6.

Fengler SA, Pearl RK. Technical considerations in the surgical treatment of colon and rectal cancer. *Seminars in Surgical Oncology* 1994;10:200-7.

Galandiuk S. Immunoscintigraphy in the surgical management of colorectal cancer. *Journal of Nuclear Medicine* 1993;34:541-4.

Gasparini M, Buraggi GL, Regalia E, Maffioli L, Balzarini L, Gennari L. Comparison of radioimmunodetection with other imaging methods in evaluating local relapses of colorectal carcinoma. *Cancer* 1994;73:846-9.

Goldenberg DM. New imaging techniques in gastrointestinal cancer. *Current Opinions in Oncology* 1993;5:697-702.

Hawes RH. New staging techniques: endoscopic ultrasound. *Cancer* 1993;71:4207-13.

Hernandez-Socorro CR, Guerra C, Hernandez-Romero J, Rey A, Lopez-Facal P, Alvarez-Santullano V. Colorectal carcinomas: diagnosis and preoperative staging by hydrocolonic sonography. *Surgery* 1995;117:609-15.

Ito K, Kato T, Tadokoro M, Ishiguchi T, Oshima M, Ishigaki T, et al. Recurrent rectal cancer and scar: differentiation with PET and MR imaging. *Radiology* 1992;182:549-52.

Kievet J, Bruinvels DJ. Detection of recurrence after surgery for colorectal cancer. European Journal of Cancer 1995;31A:1222-5.

Kroneborg O. Optimal follow-up in colorectal cancer patients: What tests and how often? *Seminars in Surgical Oncology* 1994;10:217-24.

Kronborg O, Fenger C, Deichgraeber E, Hansen L. Follow-up after radical surgery for colorectal cancer: design of a randomized study. *Scandinavian Journal of Gastroenterology* 1988; 149(Suppl):159-62.

Lai DTM, Fulham M, Stephen MS, Chu K-M, Solomon M, Thompson JF, et al. The role of whole-body positron emission tomography with [^{18}F]fluorodeoxyglucos in identifying operable colorectal cancer metastases to the liver. *Archives of Surgery* 1996;131:703-7.

Makela JT, Laitinen SO, Kairaluoma MI. Fire-year follow-up after radical surgery for colorectal cancer: results of a prospective randomized trial. Archives of Surgery 1995;130:1062-7.

Nattinger AB. Colon Cancer Screening and Detection. In Panzer RJ, Black ER, Griner PF, eds. *Diagnostic Strategies for Common Medical Problems*. American College of Physicians, Philadelphia, 1991.

Nelson RL. The decision to treat patients with recurrent colorectal cancer. *Cancer* 1993;71:4298-301.

Nelson RL. Screening of average-risk individuals for colorectal cancer and postoperative evaluation of patients with colorectal cancer. *Surgical Clinics of North America* 1996;76:35-45.

Ohlsson B, Breland U, Ekberg H, Graffner H, Tranberg K-G. Follow-up after curative surgery for colorectal carcinoma: randomized comparison with no follow-up. Diseases of the Colon and Rectum 1995;38:619-26.

Panzer RJ. Hepatic Metastases. In Panzer RJ, Black ER, Griner PF, eds. *Diagnostic Strategies for Common Medical Problems*. American College of Physicians, Philadelphia, 1991.

Petersen BM, Bass BL, Bates HR, Chandeysson PL, Harmon JW. Use of the radiolabeled murine monoclonal antibody, ^{111}In-CYT-103, in the management of colon cancer. *American Journal of Surgery* 1993;165:137-43.

Price P, Jones T, EC PET Oncology Concerted Action and EORTC PET Study Group. Can positron emission tomography (PET) be used to detect subclinical response to cancer therapy? *European Journal of Cancer* 1995;31A:1924-7.

Rafaelsen SR, Kronborg O, Larsen C, Fenger C. Intraoperative ultrasonography in detection of hepatic metastases from colorectal cancer. *Diseases of the Colon and Rectum* 1995;38:355-60.

Schiepers C, Penninckx F, De Vadder N, Mercks E, Mortelmans L, Bormans G, et al. Contribution of PET in the diagnosis of recurrent colorectal cancer: comparison with conventional imaging. *European Journal of Surgical Oncology* 1995;21:517-22.

Schlag P, Lehner B, Strauss LG, Georgi P, Herfarth C. Scar or recurrent rectal cancer: positron emission tomography is more helpful than immunoscintigraphy. *Archives of Surgery* 1989; 124:197-200.

Stark DD, Wittenberg J, Butch RJ, Ferrucci JT. Hepatic metastases: randomized, controlled comparison of detection with MR imaging and CT. Radiology 1987;165:399-406.

Strauss LG, Clorius JH, Schlag P, Lehner B, Kimmig B, Egenhart R, et al. Recurrence of colorectal tumors: PET evaluation. *Radiology* 1989;170:329-32.

Tempero M, Brand R, Holderman K, Matamoros A. New imaging techniques in colorectal cancer. Seminars in Oncology 1995;22:448-71.

US Preventive Services Task Force. Screening for colorectal cancer. In: *Guide to Clinical Preventive Services*. Second edition. Williams and Wilkins, Philadelphia, 1996.

Valk PE. Sense and sensitivity: issues in technology assessment (editorial). *Journal of Nuclear Medicine* 1996;37:1436-7.

Vitola JV, Delbeke D, Sandler MP, Campbell MG, Powers TA, Wright JK, et al. Positron emission tomography to stage suspected metastatic colorectal carcinoma to the liver. *American Journal of Surgery* 1996;171:21-6.

Yonakura Y, Benua RS, Brill AB, Som P, Yeh SDJ, Kemeny NE, et al. Increased accumulation of 2-deoxy-2-[^{18}F]fluoro-D-glucose in liver metastases from colon cancer. Journal of Nuclear Medicine 1982;23:1133-7.

VII. REFERENCES: Technical efficacy studies

Bohdiewicz PJ, Scott GC, Juni JE, Fink-Bennett D, Wilner F, Nagle C, et al. Indium-111 Oncoscint CR/OV and F-18 FDG in colorectal and ovarian carcinoma recurrences: early observations. *Clinical Nuclear Medicine* 1995;20:230-6.

Engenhart R, Kimmig BN, Strauss LG, Hover KH, Romahn J, Haberkorn U, et al. Therapy monitoring of presacral recurrences after high dose irradiation: value of PET, CT, CEA and pain score. *Strahlentherapie und Onkologie* 1992;168:203-12.

Findlay M, Young H, Dunningham D, Iveson A, Cronin B, Hickish T, et al. Noninvasive monitoring of tumor metabolism using flurordeoxyglucose and positron emission tomograph in colorectal cancer liver metastases: correlation with tumor response to fluorouracil. *Journal of Clinical Oncology* 1996; 14:700-888.

Okazumi S, Isono K, Enomoto K, Kikuchi T, Ozaki M, Yamamoto H, et al. Evaluation of liver tumors using fluorine-18-fluorodeoxyglucose PET: characterization of tumor and assessment of effect of treatment. *Journal of Nuclear Medicine* 1992;33:333-339.

Yonakura Y, Benua RS, Brill AB, Som P, Yeh SDJ, Kemeny NE, et al. Increased accumulation of 2-deoxy-2-[^{18}F]fluoro-D-glucose in liver metastases from colon cancer. Journal of Nuclear Medicine 1982;23:1133-7.

VIII. REFERENCES: Studies reviewed but not included in evidence tables

Exclusion criteria included:

- number of colorectal cancer cases < 12
- radiopharmaceutical other than FDG
- duplicated or superseded by subsequent study from the same institution.

Gupta NC, Falk PM, Frank AL, Thorson AM, Frick MP, Bowman B. Pre-operative staging of colorectal carcinoma using positron emission tomography. *Nebraska Medical Journal* 1993:30-5.

Haberkorn U, Strauss LG, Dimitrakopoulou A, Engenhart R, Oberdorfer F, Ostertag H, et al. PET studies of fluorodeoxyglucose metabolism in patients with recurrent colorectal tumors receiving radiotherapy. *Journal of Nuclear Medicine* 1991;32:1485-90.

Kim EE, Chung S-K, Haynte TP, Kim C-G, Cho B-J, Podoloff DA, et al. Differentiation of residual or recurrent tumors from post-treatment changes with F-18 FDG PET. *RadioGraphics* 1992;12:269-79.

Martin WH, Delbeke D, Patton JA, Sandler MP. Detection of malignancies with SPECT versus PET with 2-[fluorine-18]fluor-2-d-glucose. *Radiology* 1996;198:225-31.

Appendix 5

Systematic Review:
PET as a Diagnostic Test in Breast Cancer

Authors: Karen Flynn, D.D.S., M.S., *Manager, MDRC Technology Assessment Program*
Elizabeth Adams, R.R.T., M.P.H., *Management & Program Analyst, MDRC Technology Assessment Program*

Appendix 5

Systematic Review: PET as a Diagnostic Test in Breast Cancer

The final literature database searches for the systematic reviews were performed on September 10, 1996; the assessment represents peer-reviewed literature published and indexed as of that date.

This Appendix to the PET assessment presents the results of the systematic review of PET in breast cancer. A general rationale for the use of PET in oncology is supplied by Hawkins, et al. (1994) and Hoh, et al. (1994):

- many forms of cancer characteristically perturb tissue biochemical and physiological processes and PET imaging can be expected to detect the resulting abnormalities;
- reliance on tumor histology and anatomy limits the oncologist's tools for selecting optimal treatment;
- the ability to monitor metabolic responses to treatment could allow the early re-direction of therapy in patients who fail to respond to the first attempt at radiation or chemotherapy.

These and other authors (e.g., Price and Jones, 1995) report that PET studies in cancer are emerging as a major focus of the technology, both in basic research and in clinical investigations. Information gathered by the MDRC Technology Assessment Program from VA PET facilities corroborates that perception (see *Appendix 9: Experience With PET in VHA*).

Fluorine-18-fluorodeoxyglucose (FDG) is the most commonly employed radiopharmaceutical in PET cancer studies. Many neoplasms have high glycolytic rates, resulting in intracellularly trapped phosphorylated FDG that can be imaged with PET. Hawkins, et al. (1994), note that tumor-specific biochemical characteristics of glucose transport and phosphorylation may affect quantitative estimates of tumor glucose metabolism with FDG PET, and that investigations are under way to define these characteristics. However, these uncertainties may be of less concern with qualitative or semiquantitative FDG PET cancer studies because the primary intent of such studies is to detect and map tumor foci, not to rigorously quantify tumor glycolytic rates.

In some instances, PET imaging techniques have been modified to meet the needs of cancer diagnosis. Most PET systems allow axial fields of view (the length of the body encompassed by a series of cross sectional images) of approximately 10 cm. Cancer is frequently distributed beyond this field of view, and whole body image acquisition procedures have been developed (Hoh, et al., 1993). Since it is impractical to apply standard transmission scanning attenuation correction methods to these procedures, whole body PET imaging is primarily useful as a qualitative indicator of disease distribution.

Nieweg (1994) and Price and Jones (1995) define a number of potential applications for PET in oncology. These include:

- tumor detection (although PET images offer insufficient structural detail and should not be used to visualize anatomy; registration techniques to combine PET and anatomic imaging into a single image are under development to circumvent this limitation);

- staging (particularly using whole-body imaging methods) although there is a lower limit to the size of metastases that can be detected by PET;

- detection of local recurrence of disease, since anatomically-based imaging is often limited by the effects of treatment;

- prediction of tumor response to chemotherapy;

- treatment monitoring.

I. BACKGROUND

A. General sources

The discussion in this overview section, unless otherwise noted, is based on information distributed by the National Cancer Institute (NCI) in September, 1996 through its on-line Physician Data Query (PDQ) system.

B. Description

Breast cancer is the most common life-threatening malignancy among adult women in every major ethnic group in the United States (Kelsey and Horn-Ross, 1993). It is highly treatable by surgery, radiotherapy, chemotherapy, and hormonal therapy, and can be cured when detected in its early stages. Accordingly, screening programs using mammography and/or breast physical examination have been widely implemented and tested for efficacy.

C. Epidemiology

The American Cancer Society (Cancer Facts and Figures, 1996) estimates that 184,300 new cases will be diagnosed in 1996; approximately 1 of every 8 women (12.5%) will develop breast cancer during the course of a maximum life expectancy. A 4% per year increase in the incidence rate has been observed between 1982 and 1987; part of the increase may be attributable to screening programs which detect some breast cancers while they are still non-invasive (carcinoma in situ) and before they become clinically apparent (Liff, et al., 1991).

In 1992, there were 1,199,370 women veterans (4.4% of the total veteran population) (Department of Veterans Affairs, National Survey of Veterans, 1995). An HSR&D service-directed research project (Feussner and Hynes, abstract in: *Projects Receiving Funding in Fiscal Year 1995*) found an urgent need to promote early breast cancer detection program for women veterans and to further assess their risks for breast cancer.

Although the incidence of breast cancer is increasing, early detection by means of routine mammography examination and improved treatment have allowed the death rate to remain essentially unchanged since 1930. Breast cancer is now the second most common cause of cancer death among women (after lung cancer). Approximately 44,300 women and 260 men will die of breast cancer in the United States in 1996; 15% of these deaths will occur among premenopausal women, making breast cancer a leading cause of years of potential life lost.

The risk of breast cancer increases with age in high-risk areas such as the United States, but the increase with age is less steep after about age 45-50 than it is during the reproductive years. The specific factors associated with a three- to four-fold relative risk for the development of breast cancer include; first-degree relative (mother or sister) with breast cancer; prior breast cancer; nulliparity; first childbirth after age 30; early menarche or late menopause; and radiation exposure, particularly in the prepubertal years. Other factors implicated in increasing the risk for breast cancer include: hyperplastic fibrocystic disease with atypical epithelial cells; use of oral contraceptives by young women before a first pregnancy; long-term use of non-conjugated estrogens; and the fat content of the diet.

The etiology of breast cancer remains undefined. Some of the risks listed above may relate to etiology; genetic factors, hyper- or un- opposed estrogen activity over a long reproductive life span, and some dietary factors are thought to contribute.

D. Diagnosis

Invasive carcinomas of the breast can arise from both the lobular and ductal components of breast tissue. Approximately 85% of invasive carcinomas are infiltrating ductal carcinomas; 10% are infiltrating lobular carcinomas; and the remaining 5% include medullary, mucinous, tubular, and adenoid cystic histopathological varieties. Although use of mammography has reduced the average size of newly diagnosed breast lesions to 2.5 cm, many breast cancers are still discovered by patients or physicians during routine examinations (and are therefore clinically detectable at diagnosis). Currently, between 30% and 50% of breast cancer lesions have progressed to involve the axillary lymph nodes at diagnosis.

Following initial diagnosis, that diagnosis is confirmed, stage of disease is evaluated, and therapy is selected. Diagnosis may be confirmed by aspiration cytology, solid core needle biopsy, or incisional or excisional biopsy. When technically possible, an excisional biopsy both supplies material for definitive diagnosis of a breast lesion, and acts as the primary surgical treatment (which also includes axillary dissection) for a woman with a small primary tumor who will subsequently receive radiation therapy as her principal adjuvant therapy. Additional prognostic factors determined during the diagnostic process include stage of the disease, histologic and nuclear grade, and hormone receptor status.

E. Staging, treatment, and survival

The staging system for breast cancer provides a strategy for grouping patients with respect to prognosis. Both survival and risk of relapse after treatment are associated with the degree of progression of the disease at diagnosis. The size of the primary tumor, the presence or absence of histologically confirmed lymph node involvement, the number of nodes involved, and the metastatic spread of the disease at diagnosis and initial surgical treatment are used in staging (tumor, node, metastases, or TNM staging). Stage-specific survival rates have increased only slightly since the 1970s.

Breast cancer is a common cause of death among women, and a significant contributor to potential years of life lost. In order to reduce breast cancer mortality, a number of approaches could be taken. Since the cause of breast cancer remains elusive for the majority of women with the disease, primary prevention is not yet feasible. Screening programs attempt to identify early breast cancers when they are highly curable. Mammography (with/without clinical breast exam) is currently the best screening tool available, and has been shown to be of value in randomized trials. Combined Swedish trial data indicate that mammography produced an overall reduction in breast cancer mortality of 29% during 12 years of follow up in women over 50 and a 13% reduction in younger women (Blamey, et al., 1994), although screening in unselected populations of women under 50 remains controversial.

Finally, treatment for more advanced breast cancer can be improved. Currently, treatment is associated with less than optimal outcomes with respect to both survival and quality of life during treatment, and is an area of intense research activity. The NCI notes that, even when standard therapy is effective, patients with breast cancer are appropriately considered as candidates for clinical trials designed to improve therapeutic results and to decrease the morbidity of treatment.

Therapeutic decisions are made in part according to staging categories, but primarily according to lymph node status, estrogen and progesterone receptor levels in the tumor, menopausal status, and the overall health of the patient. In general, initial treatment for breast cancer is surgical removal of the tumor, followed by radiation therapy to enhance loco-regional control and/or by chemotherapy to reduce the risk of recurrence due to micrometastases or to control identified metastases. Table 1 provides information on the

TNM staging system in breast cancer, on the therapeutic options at each stage, and on survival.

The lifetime cost of treating breast cancer has been estimated at $61,000 in 1991 dollars (over $10 billion in the U.S. in 1991). These costs are only those borne by the medical care system; additional unquantified costs are associated with pain, suffering, and anxiety to patients and their families. The U.S. General Accounting Office has concluded that the best prospect for reducing breast cancer mortality is through increased utilization of screening mammography. Cost-effectiveness studies estimate that mammography has a marginal cost of approximately $36,000 (over clinical breast exam alone) per year of life saved, which is comparable to the cost effectiveness of treating hypertension and hypercholesterolemia (White, et al., 1993).

F. Potential roles for PET

Screening strategies using mammography and breast physical examination are widely available. Mammography is relatively expensive (on a population basis), requires high levels of technical expertise, and detects only 95% of breast cancers but has been validated in large randomized trials (Blamey, et al., 1994; Morrison, 1993). The number of women receiving mammography is increasing: the percentage of women over 40 years of age who had ever obtained at least one mammogram rose from 38% in 1987 to 60% in 1990, and the percentage of women who had a mammogram in the previous year rose from 17% to 33%.

PET researchers (e.g., Adler, 1993) acknowledge that the high cost (and limited availability) of PET argue against its potential to replace mammography for screening large populations of asymptomatic women. However, other potential applications of PET in the management of patients with breast cancer have been defined in the literature:

- screening in subgroups of women, e.g., those with breast implants (Wahl, et al., 1994), and those with prior breast radiotherapy, multiple breast masses and history of negative biopsy results, or severe fibrocystic disease (Tse, et al., 1992);

- monitoring response to chemotherapy (Wahl, et al., 1993; Tse, et al., 1992);

- nonsurgical evaluation of breast disease (Tse, et al., 1992; Adler, et al., 1993; Avril, et al., 1996b; Scheidhauer, et al., 1996);

- selection of patients for axillary dissection and for preoperative chemotherapy (Tse, et al., 1992; Adler, et al., 1993; Avril, et al., 1996a; Scheidhauer, et al., 1996)

- quantification of tumor glycolytic rate as a prognostic factor (Tse, et al., 1992; Avril, et al., 1996b).

II. RESULTS

Twenty-three articles from MEDLINE and other database searches and from the bibliographies of initially retrieved articles were selected as meeting the screening criteria. After review, 13 (56%) met inclusion criteria for assignment to the following levels of the diagnostic efficacy hierarchy (Fryback and Thornbury, 1991; *Appendix 2: Assessing Diagnostic Technologies*): 8 met the definition of technical efficacy (See reference list; full data abstraction tables for technical efficacy studies are on file with the MDRC Technology Assessment Program) and 5 met most or all of the evidence-based criteria for studies of diagnostic accuracy (Table 3). Table 2 summarizes cross-study findings on PET and alternative technologies. These studies have methodologic limitations, and accordingly were not combined statistically for this review.

The potential uses of PET in breast cancer outlined above have been addressed only anecdotally or in studies that should be considered preliminary. Subgroup screening has been studied in two patients with breast implants (Wahl, et al., 1994) and in 3 patients with radiodense breasts (in a series of 14 reported by Tse, et al., 1992). Changes in tumor glucose metabolism during chemotherapy have been reported in small series of patients (Bassa, et al., 1996; Bruce, et al., 1995; Jansson, et al., 1995; Mortimer, et al., 1996; Wahl, et al., 1993). Correlation of tumor glycolytic rate to other prognostic features were studied by Avril, et al., (1996b) with no association established.

Studies presented below addressed the non-invasive determination of primary breast disease and axillary lymph node involvement using FDG PET. One small case-control study was identified (Nieweg, et al., 1993) and was included, in spite of its small size. The remaining studies were series of patients who presented for surgical evaluation of breast masses; patients without disease within the series were internal controls used for comparison. Patients evaluated for axillary node involvement were accrued from those with malignant primary breast disease, and patients with benign nodal disease served as internal controls.

All patients in these case series had suspected or biopsy-proven breast cancer, and relatively low proportions of patients had benign lesions. Therefore, the pre-test probability of disease in the study populations was high. Variations in reporting across studies with respect to unit of analysis (by patient, lesion or axillae) and extent of disease (size of primary tumor and level of axillary lymph node involvement) may further affect the generalizability of these results. All studies provided information on the comprehensiveness of blinding of the test interpreters to the gold standard. However, none of these studies met strict evidence-based criteria for blinding, because determination of the diagnostic gold standard independent of the PET results could not be ascertained. This will likely result in significant bias and inflated estimates of accuracy. These results should be interpreted accordingly.

A. Defining primary breast disease

Adler, et al., (1993) and Avril, et al., (1996b) assessed the ability of PET to detect primary breast disease both quantitatively and qualitatively. Nieweg, et al., (1993b) used a case control design, but failed to report information on criteria for judging images positive or negative for cancer and used only one image interpreter. No data on alternative technologies were presented. Variations in blinding of the image readers to either the gold standard or to other clinical information and in choice of a retrospectively determined cut-off used in the quantitative analyses occurred across these studies. Only Scheidhauer, et al., (1996) reported operating characteristics for alternatives to PET, but the evaluation of clinical examination and mammography was not described with sufficient detail to be reproducible. These studies have limitations in reporting and study design and should be considered preliminary.

B. Defining axillary lymph node involvement

Adler, et al., (1993) and Scheidhauer, et al., (1996) assessed a small subset of patients with malignant primary breast cancer for axillary lymph node involvement but reported no data on alternative technologies. Avril, et al., (1996a) presented results in a larger group of patients combined with results from 10 patients with benign primary breast disease to compare the accuracy of PET to clinical exam. Subgroup analyses according to primary tumor stage were conducted for PET but not for clinical examination. Results were presented as point estimates with 95% confidence intervals, which ranged widely due to small study size. These results should be interpreted with caution.

C. Detecting distant metastases

Scheidhauer, et al., (1996) also reported anecdotal evidence on a subgroup of 8 patients (23 total lesions) with distant metastases at the time of diagnosis. Small study size and limitations in study design suggest that these results should be considered preliminary.

III. SUMMARY

Preliminary studies into the role of FDG PET in the diagnosis and management of breast cancer have been reported in the literature. Table 2 summarizes findings from these studies on the diagnostic accuracy efficacy of PET. These studies received low methodology grades because of limitations in blinding, small study size, and incomplete reporting. The MDRC Technology Assessment Program did not identify any published studies that documented PET imaging at higher levels of the diagnostic efficacy hierarchy.

Only one study (Nieweg, et al., 1993) used a case control design; the remaining studies were uncontrolled or used small numbers of internal controls. None of the studies met strict evidence-based criteria for evaluation of diagnostic tests. The prevalence of malignancy in these study populations is high, may be weighted toward severe disease, and may not provide accurate estimates of accuracy. Predictive values should be interpreted accordingly. These data are further complicated by variations in reporting with respect to the unit of analysis (by patient, lesion, or axillae) and extent of disease (primary tumor size and level of axillary node involvement), and may not be generalizable to a population of mammographically tested patients with a lower prevalence of malignancy. All authors stressed the preliminary nature of these results and recommend assessment of PET in larger trials.

IV. DISCUSSION

The studies in Tables 2 and 3 report accuracy findings for detecting axillary lymph node involvement. Axillary dissection with histopathology of dissected nodes supplies information critical to subsequent treatment decisions, is currently recommended by the NCI for most patients with Stage 1 or higher disease, but is associated with significant morbidity. Under investigation are less invasive surgical methods and improved imaging modalities, including PET, which are used to map axillary lymph node involvement. However, selecting patients for axillary dissection based on PET studies would be extremely premature given the currently available PET research data. Published PET data are based on very small numbers of patients (compared to the tens of thousands who have enrolled in studies of screening and treatment options) and should be confirmed in larger, more rigorous studies before being incorporated into clinical practice. The efficacy of chemotherapy in patients with positive nodes has been demonstrated in large, randomized trials; inappropriate assignment of patients to no chemotherapy (or to unneccessary exposure to the morbidity of chemotherapy) in the absence of axillary dissection should be avoided.

Future PET research in breast cancer is likely to involve improvements to resolution and diagnostic performance. High resolution positron emission mammography is under development (Thompson, et al., 1995), Although they do not report using the technology in patients, these authors hypothesize that this modification to existing PET technology will be an adjunct to conventional mammography and may eventually be an alternative to needle or surgical biopsy.

A. Alternatives to PET in some of its potential breast cancer applications

As noted above, improvements to breast cancer imaging and patient management are areas of intense research activity. Mammography in asymptomatic women is associated with a high sensitivity but a high yield of false positive results, a low positive predictive value, and a low specificity. However, technical improvements, increasing experience of radiologists who specialize in screening, and quality assurance criteria are increasing the accuracy of mammography and the diagnostic yield of subsequent procedures in established, high quality screening programs (Tubiana, et al., 1994).

Adler and Wahl (1995) review new methods for imaging the breast which may correct some of the shortcomings of mammography. These methods address both detection and classification of breast lesions, and include MRI of the breast, digital mammography, computer aided diagnosis, SPECT, and PET. These authors confirm that all of these methods would require significant technical refinement before replacing mammography for the detection of breast cancers.

The classification of mammographically detected lesions as benign or malignant (i.e. improving the specificity of mammography) is a particularly fertile area for continued research. As many as 70% and 85% of women who receive biopsies based on suspicious mammographic findings have benign conditions (Parker, et al., 1995). A non- or less invasive method for evaluating breast disease prior to biopsy which would reduce the number of surgical biopsies for benign lesions is desirable.

According to a review by Harms, et al., (1994), studies using MRI of the breast have demonstrated consistent contrast enhancement of malignant lesions and the lack of contrast enhancement of benign conditions. MRI used before surgical biopsy has been demonstrated to have a high negative predictive value and the ability to reduce the number of biopsies performed for benign lesions. Technical advances, such as dynamic contrast imaging, which will improve the diagnostic performance characteristics of breast MRI, are under investigation.

Biopsy techniques that are less invasive than open surgical biopsy, and that would reduce the adverse effects of unnecessary open biopsy procedures on health care resources and patients' psychosocial well-being, have been developed and have diffused into many health care settings. Stereotactic mammography to localize suspicious breast lesions, followed by fine needle aspiration or large needle core biopsy, are among these techniques. The equipment for these procedures is marketed in the United States. An analysis of over 6000 solid core biopsy results in the United States has recently been published; the authors concluded that the procedure is a reproducible and reliable alternative to surgical biopsy (Parker, et al., 1994). Tubiana, et al., (1994) report that fine needle aspiration cytology, in combination with increasing radiological expertise, can lead to malignant/benign biopsy ratios in specialist centers of between 3/1 and 10/1, minimizing the number of unnecessary benign biopsies.

B. A breast cancer research agenda

Given the population impact of breast cancer, it is reasonable to consider both new and existing screening, diagnostic (including PET), and treatment technologies within the context of the overall knowledge base and a strategic research agenda. Such an approach could provide a framework for evaluating the technologies' population and societal impacts.

The Institute of Medicine (IOM), in a 1993 report to the U.S. Army Medical Research and Development Unit on strategies for managing the breast cancer research program, lists the deficiencies in the current knowledge base on breast cancer. These include: biologically based treatments directed at precise targets are tantalizingly possible, but require major continuing research efforts; no dominant etiology for breast cancer has emerged from extensive epidemiologic studies, making it unlikely that quick and easy prevention strategies can be implemented; access to screening and treatment is problematic for minority women; the currently available treatments are less than completely effective and exact a substantial physical and emotional toll on the women who receive them; breast cancer is a very heterogeneous disease and it remains difficult to determine with any certainty the best therapeutic regimen for any particular woman; and far too few data are available to assess the effectiveness of various therapeutic interventions or how best to deliver them.

In response to the deficiencies in the knowledge base listed above, the IOM developed the following questions to guide research:

- What genetic alterations are involved in the origin and progression of breast cancer?

- What are the changes in cellular and molecular functions that account for the development and progression of breast cancer?

- How can endogenous and exogenous risk factors for breast cancer be explained at the molecular level?

- How can investigators use what is known about the genetic and cellular changes in breast cancer to improve detection, diagnosis, prevention, treatment, and follow up?

- What is the impact of risk, disease, treatment, and ongoing care on the psychosocial and clinical outcomes of breast cancer patients and their families?

- How can investigators define and identify techniques for delivering effective and cost-effective health care to all women to prevent, detect, diagnose, treat, and facilitate recovery from breast cancer?

The IOM felt that these questions offered a usable framework for defining the goals of a research program and evaluating the relevance of proposed research.

V. SUGGESTIONS FOR FURTHER RESEARCH

Breast cancer offers a somewhat different context for an evaluation of the potential roles of PET than do some of the other oncology indications for PET that are addressed in this report. Breast cancer makes a substantially larger contribution to potential years of productive life lost in the general population than do the other malignancies, and consumes a proportionate amount of health care resources. Mammography is among the only screening tests in routine use that is supported by evidence from large randomized clinical trials. Finally, a greater number of competing technologies (including those that improve the results of mammography) are under development.

Veterans Health Administration breast cancer research efforts could focus on questions similar to those of the IOM. Proposals to apply PET imaging to clinical management of breast cancer patients could then be evaluated within that framework.

Table 1 Breast cancer staging, treatment options, and survival

Stage	***Pathologic description***		***Treatment options***	***5-year survival***
0	carcinoma in situ • nonpalpable lesions discovered on screening mammography • accounts for 15 - 20% of all breast cancers	intraductal • presents as mammographic microcalcifications or a soft tissue abnormality	• mastectomy with excision of lymph nodes around axillary tail of breast (but without a formal axillary dissection) results in a local and distant recurrence rate of 1 - 2% • conservative surgery with radiotherapy results in recurrence rate of 9 - 21% • salvage of recurrences is feasible, with survival comparable to upfront mastectomy	> 95%
		lobular • generally widely distributed throughout breast and frequently bilateral • indicates 25% chance of developing invasive cancer within 25 years	• clinical management controversial; options include no treatment after biopsy with careful follow up, or bilateral prophylactic mastectomy • axillary dissection not necessary • patients who have undergone local excision are eligible for a large multicenter RCT of tamoxifen to prevent development of invasive cancer	
I	- tumor ≤ 2.0 cm in greatest diameter - no regional lymph node metastases - no distant metastases		• 21% of patients managed with surgery alone may ultimately relapse • breast conserving surgery followed by radiation therapy provides tumor control equivalent to more extensive surgical procedures; approximately 20% of patients experience local recurrence • axillary dissection should be performed • suitable ER negative patients receive adjuvant chemotherapy with a proven effective regimen; a group of patients with small tumors who do not benefit from adjuvant therapy may be identified • ER positive patients receive adjuvant tamoxifen	85%
II	• tumor > 2cm but ≤ 5 cm in greatest diameter • no nodal metastases, or metastases to movable ipsilateral axillary nodes • no distant metastases	negative nodes	• survival is equivalent with any of the surgical options (mastectomy, mastectomy with reconstruction, or conservative surgery plus radiation therapy) • radiation therapy to the chest wall and regional nodes should be considered for patients at high risk of local recurrence (with known residual disease) • adjuvant combination chemotherapy with a proven effective regimen for both pre- and post-menopausal ER negative patients • adjuvant tamoxifen with established schedule for ER positive patients • optimal adjuvant therapy has not been defined for any subset of patients; all patients and their physicians are strongly encouraged to participate in controlled clinical trials	66%
		positive nodes	• survival is equivalent with any of the surgical options (mastectomy, mastectomy with reconstruction, or conservative surgery plus radiation therapy) • radiation therapy to the chest wall and regional nodes should be considered for patients at high risk of local recurrence (with known residual disease or 4 or more involved nodes) • adjuvant combination chemotherapy with a proven effective regimen for both pre- and post-menopausal patients • tamoxifen, either alone or combined with chemotherapy, prolongs DFS when administered for 24 months to postmenopausal women with axillary node metastases • optimal adjuvant therapy has not been defined for any subset of patients; all patients and their physicians strongly encouraged to participate in controlled clinical trials	

Stage	*Pathologic description*		*Treatment options*	*5-year survival*
III	• tumor any size • metastases to movable ipsilateral axillary nodes or metastases to ipsilateral nodes fixed to one another or to other structures • no distant metastases	operable	• modified radical mastectomy or radical mastectomy • radiation therapy • combination chemotherapy with or without hormones is given in conjunction with surgical and radiotherapeutic procedures • tamoxifen as postoperative adjuvant hormonal therapy for postmenopausal patients with high ER and PR levels is under investigation	41%
		inoperable	• newly diagnosed patients should be considered candidates for one of the clinical trials in progress to improve therapeutic results • incisional biopsy followed by radiation to primary tumor and regional lymph nodes • combination adjuvant chemotherapy • if chemotherapy is contraindicated, hormonal therapies may be used for patients whose tumors are ER and PR positive after surgery and radiotherapy • initial chemotherapy followed by surgery and/or radiation therapy is under investigation • phase II studies evaluating new chemotherapeutic or biologic agents may be considered for patients whose local disease is not controllable by other means	
IV	• tumor any size • any pattern of axillary lymph node metastases • distant metastases present		• all patients with stage IV disease should be considered candidates for one of the ongoing clinical trials to improve therapeutic results • surgical biopsy • in patients with ER and PR positive tumors and no visceral disease, hormonal therapy is used • in patients with ER and PR negative tumors or visceral disease, combination chemotherapy is used	10%
Recurrent	localized recurrence after breast conserving surgery	• subset of good prognosis patients with locally recurrent disease in the breast at 1-9 years post lumpectomy plus radiotherapy	• salvage mastectomy plus radiotherapy	65-80%
	other/widespread recurrence	• visceral disease absent, ER and PR positive or unknown, disease free interval > 2 years	• hormonal therapy/oophorectomy for premenopausal patients • antiestrogen/progesterone therapy with tamoxifen for postmenopausal patients	treatment is palliative
		• recurrence localized or visceral	• surgery and/or radiotherapy	treatment is palliative
		• relapse after response to additive hormonal therapy	• other forms of hormonal therapy not previously used	treatment is palliative
		• visceral disease, ER and PR negative, or disease free interval < 2 years	• combination chemotherapy	treatment is palliative

Abbreviations:
RCT, randomized controlled trial
DFS, disease free survival
ER, estrogen receptor
PR, progesterone receptor

Table 2 Summary of the Literature: Diagnostic accuracy efficacy studies of PET and alternatives in breast cancer

Notes: All studies except Nieweg, et al., (1993b), which was a case-control study, were series of patients presenting for surgical evaluation of breast masses (a high index of suspicion of malignant disease) and included internal controls as the comparison group. Predictive values should be viewed accordingly. Studies assessing axillary node involvement included patients with malignant primary breast disease. Results from Avril, et al., (1996b) were reported as ranges of data from all subgroup analyses. Results from Avril, et al., (1996a) included all patients with benign and malignant primary disease and represent 95% confidence intervals; subgroup analyses were not reported because of their small study size. None of these studies met strict evidence-based medicine criteria for blinding, but all studies provided data on the comprehensiveness of blinding of test interpreters to the gold standard.

Where substantial duplication in purpose of study, patients studied, and results in multiple studies from the same institution could be inferred, only the most recent, largest, most rigorously designed, or most comprehensive was included in the table. Although data from both studies by Avril and associates (1996a and 1996b) represent the same patient population, these studies addressed different purposes; inclusion of both publications was felt to be warranted.

Abbreviations are listed at the end of the table.

Role *(Note: some studies assessed multiple roled)*	**Study**	**N**	**Operating Characteristics***			**Evidence-Based Medicine Criteria****			**Methodologic Quality Grade*****
			PET	**Clinical Exam**	**Mammography**	**comparison group**	**histologic gold standard**	**blinding**	
Defining primary disease	Adler, et al., 1993	27 positive lesions 8 negative lesions	Se=96% Sp=100%			+ internal	+	+	C
	Nieweg, et al., 1993b	11 cases 8 controls	Se=91% Sp=100%			+	+	+	C
	Avril, et al., 1996b	41 positive lesions 31 negative lesions	Se=68%-94% Sp=84%-100% PPV=87%-97% NPV=70%-93%			+ internal	+	partial	D
	Scheidhauer, et al., 1996	23 malignant cases 7 benign cases	Se=91% Sp=86%	Se=74% Sp=71%	Se=86%	+ internal	+	partial	D
Defining axillary node involvement	Adler, et al., 1993	9 positive axillae 10 negative axillae	Se=90% Sp=100%			+ internal	+	+	C
	Avril, et al., 1996a	24 positive axillae 27 negative axillae	Se=57%-93% Sp=81%-100% PPV=75%-100% NPV=66%-100%	Se=36%-78% Sp=66%-96% PPV=30%-70% NPV=51%-85%		+ internal	+	+	C
	Scheidhauer, et al., 1996	9 malignant cases 9 benign cases	Se=100% Sp=89%			+ internal	+	partial	D
Detecting distant metastases	Scheidhauer, et al., 1996	8 positive lesions 15 negative lesions	Se=100% Sp=100%			+ internal	+	partial	D

N, number of study subjects included in analysis; unless otherwise noted, data are analyzed by subject
Se, sensitivity
Sp, specificity
PPV, positive predictive value
NPV, negative predictive value

* operating characteristics defined in *Appendix 2: Assessing Diagnostic Technologies, page 5-7*
** *Appendix 2, page 8*
^^^ *Appendix 2, page 9*

Table 3 Diagnostic efficacy of FDG PET and alternatives in breast cancer

Note: All of the studies in this table met most of the evidence-based medicine criteria for diagnostic test evaluations. Nieweg, et al., (1993b) was included (in spite of its small number of breast cancer cases) because it is a case-control study. All other studies are case series with internal controls (i.e., patients with benign primary masses or benign axillary nodes), and it was possible to calculate sensitivity and specificity for PET in those studies; however, all of these patients had suspected breast cancer and relatively low proportions of patients had benign lesions, making the pre-test probability of disease in the study populations very high. Accordingly, predictive values were only reported for both studies by Avril and associates (1996a and 1996b), because there was roughly an equivalent proportion of malignant and benign disease. Studies assessing axillary node involvement included subsets of only those patients with malignant primary breast disease from corresponding case series.

Where substantial duplication in purpose of study, patients studied, and results in multiple studies from the same institution could be inferred, only the most recent, largest, most rigorously designed, or most comprehensive was included in the table. Although data from both studies by Avril and associates (1996a and 1996b) are derived from the same patient population, these studies addressed different purposes, and inclusion of both was felt to be warranted.

All studies in the table compared PET to the "gold standard" of histopathology of surgical specimens, which is the reference test for the operating characteristics reported in the "Results/Comments" column.

Abbreviations are listed at the end of the table.

Study	***Patients/Methods***	***Results/Comments***
Adler, et al., 1993 *University Hospitals of Cleveland*	**Purpose** • to determine the relationship between FDG uptake and tumor histology in patients with newly discovered breast masses • to determine the ability of PET to detect axillary node metastases prior to dissection **Cases** 28 patients with newly discovered breast masses ≥ 1 cm in diameter • 35 breast lesions (27 malignant, 8 benign) • 19 axillae from 18 patients (9 malignant, 10 benign) • first consenting patients who could be scheduled for PET prior to biopsy **Methods** • all subjects received breast PET scans after 4 hour fasts • 25/28 had surgical biopsy • 3/28 had fine needle aspiration • 19 patients diagnosed with breast cancer subsequently underwent PET scanning of axillary region and axillary dissection • PET results compared to histopathologic findings • PET image readers were blinded to pathologic and axillary dissection findings, but were not blinded to physical exam or mammography results • single consensus visual/qualitative interpretation (5 point scale, with 4 or 5 scored as malignant) • quantitative (DUR) measures of FDG uptake obtained with ROIs **Study design limitations** • numbers of cases and internal controls not equivalent (high prevalence of malignancy) • blinding of readers to gold standard, but not to exam or mammography • comprehensiveness of nodal sampling not indicated	**Quantitative analyses** • benign cysts (mean DUR = 0.2 ± 0.3) had diminished FDG uptake • solid benign lesions, mean DUR = 1.8 ± 0.5 • malignancies, mean DUR = 12.8 ± 9.3 • cut point DUR = 2.4 retrospectively found to discriminate benign vs malignant lesions (p = .0005) • cut point DUR = 1.0 retrospectively found to discriminate between cystic and solid benign lesions • within malignancies, DUR was significantly correlated with histologic grade (Spearman rho = .55, p = .006) **Diagnosing primary disease (27 malignant, 8 benign)** Se = 96%; Sp = 100% (based on qualitative scores) **Axillary node involvement (9 cases, 10 controls)** Se = 90%; Sp = 100% (based on qualitative scores) **Authors' comments** • high cost means that PET is unlikely to be used in screening • may eliminate need for biopsy in patients with mammographically indeterminate or occult lesions • patients with high pre-biopsy probability of disease (based on PET) might be spared a diagnostic procedure and proceed directly to definitive therapy • PET may impact management by indicating which patients have axillary nodal involvement • high specificity cannot be generalized to population of all women with breast masses ≥ 1 cm (due to high prevalence of malignant disease in sample)

Study	Patients/Methods	Results/Comments
Nieweg, et al., 1993b *M.D. Anderson Cancer Center, University of Texas*	**Purpose** to investigate the sensitivity and specificity of FDG in the detection of breast cancer **Cases** • 11 patients with biopsy-confirmed breast cancer (including one with both cancer and fibrocystic disease) • median size of tumors was 2.5 cm **Controls** 8 subjects without cancer (including 3 with previous mastectomy but currently disease free, one with 2 cysts, 3 healthy volunteers, fibrocystic lesion from one of cases) **Methods** • all subjects scanned after 4 hour fasts • TNT ratios calculated using contralateral breast and ROIs • one image reader who was blinded to clinical findings **Limitations of study design** • comprehensiveness of nodal sampling not indicated • qualitative image interpretation implicit, but criteria for judgment re presence of disease not given	**Diagnosing primary disease** • *Se = 91%; *Sp = 100% • single false negative in patient with 1 cm tubular carcinoma associated with focus of invasive ductal carcinoma • TNT ratios of tumors from 1.0 to 15.3 (median 4.9) **Axillary node involvement** • due to limitations in axial field of view, only 5 patients evaluated for axillary involvement with PET • TNT ratios for nodes from 2.1 to 29.4 **Authors' comments** • PET images easy to evaluate: no equivocal results, myocardial uptake did not interfere with image interpretation • no adverse reactions to FDG • unclear whether TNT ratios are adequate for monitoring response to chemotherapy
Avril, et al., 1996a *Technische Universität, Munich, Germany*	**Purpose** to evaluate preoperatively the diagnostic accuracy of PET for detecting axillary lymph node metastases in women with suspected breast cancer **Cases** 51 women with newly discovered breast tumors who were scheduled to undergo surgery • primary disease: 41 malignant, 10 benign • axillary lymph node status assessed in patients with malignant primary tumor, excluding 4 with locally advanced disease: 24 malignant, 17 benign **Methods** • all patients fasted for at least 4 hours before PET • SUVs calculated • images analyzed qualitatively by two independent interpreters blinded to other diagnostic tests • criterion for positive PET study: foci of increased FDG uptake reached by consensus • criterion for positive clinical exam: detection of enlarged, palpable lymph nodes or conglomerate masses in the axilla • axillary lymph node status confirmed by histopathology • PET and CE compared to histopathology **Limitations of study design** • number of cases and internal controls not equivalent (high prevalence of malignancy) • comprehensiveness of nodal sampling unclear • independence of test result and determination of final diagnosis unclear	**Detecting axillary node involvement (24 malignant, 27 benign)** *overall (reported with 95% CI)* PET: Se=79% (57%-93%); Sp=96% (81%-100%); PPV=95% (75%-100%); NPV= 84% (66%-95%) Clinical Exam: Se=58% (36%-78%); Sp=85% (66%-96%); PPV=78% (30%-70%); NPV= 70% (51%-85%) • includes 10 with benign primary disease **Detecting axillary node involvement (6 malignant, 12 benign)** *Stage pT1 only (reported with 95% CI)* PET: Se=33% (4%-78%); Sp=100% (73%-100%); PPV=100% (15%-100%); NPV= 75% (47%-93%) **Detecting axillary node involvement (18 malignant, 5 benign)** *Stage > pT1 (reported with 95% CI)* PET: Se=94% (72%-100%); Sp= 100% (47%-100%); PPV=100% (80%-100%); NPV= 83% (35%-100%) **Authors' comments** • smallest tumor-infiltrated lymph node visualized by PET was 0.8 cm in diameter • sensitivity of PET imaging of axillary lymph node depends on extent of lymph node involvement (unable to verify due to small numbers) • spatial resolution and partial volume effect limit assessment of small axillary lymph nodes • further studies are needed to compare accuracy and cost-effectiveness of PET with other imaging methods available for staging procedures • authors reported an increased sensitivity of PET with an increased number of involved lymph nodes

Study	***Patients/Methods***	***Results/Comments***
Avril, et al., 1996b *Technische Universität München, Munich, Germany*	**Purpose** • to evaluate the ability of PET to differentiate malignant versus benign breast tumors using visual and quantitative analysis • to compare regional FDG uptake in breast cancer with histology • to correlate FDG uptake with grade and size of tumors, estrogen receptor (ER)/progesterone (PR) status and rate of cell proliferation **Cases** 51 women with 72 proven breast lesions (41 malignant, 31 benign) who presented with abnormal mammography or palpable breast lesions and who were scheduled to undergo surgery • patients with prior breast surgery within the last 3 months or who had undergone chemotherapy or radiation therapy • size of breast lesions ranged from 0.3 cm to 9.0 cm (mean diameter= 2.5 cm ± 1.8 cm **Methods** • patients fasted for at least 4 hours before PET • all patients studied in prone position • ROIs over all histologically proven breast lesions identified; for lesions that could not be clearly identified by increased FDG uptake, surgeon's report used to position ROI • SUVs calculated for all histologically confirmed breast tumors; partial volume correction calculated • PET visual analysis reached by consensus performed by two observers blinded to clinical history, examinations, and histology; regional FDG uptake classified as unlikely, probable, and definite • ROC curves, correlations of FDG uptake and tumor size, blood glucose level, cell proliferation, histopathologic grading ER and PR performed only in tumors > 1 cm in size **Limitations of study design** • number of cases and internal controls not equivalent (high prevalence of malignancy) • partial blinding of PET readers to data in surgeon's report used for anatomical positioning	**Diagnosing primary disease using visual analysis** *including all lesions regarded as definite and probable malignant* Se=83%; Sp=84%; PPV=87%; NPV=87% *including only those lesions with definite malignant findings* Se= 68%; Sp=97%; PPV=97%; NPV=70% *including lesions >1 cm regarded as definite and probable malignant* Se=94%; Sp=84%; PPV=87%; NPV=93% *including lesions > 1 cm regarded as definite malignant findings* Se=78%; Sp=97%; PPV=97%; NPV=79% **Quantitative analysis** *expressed as mean ± SD* malignant = 3.3 ± 1.8 vs. benign= 1.4 ± 0.5 ($P < .01$) *from ROC analysis, using threshold SUV of 2.5* Se=75%; Sp=100% *from ROC analysis, using threshold SUV of 2.5*, with partial volume correction Se=92%; Sp=97% • differences between corrected and uncorrected SUV values not statistically significant **Other findings** • the reproducibility of ROI positioning in a subset of 20 patients was elevated; interobserver variability was r= .91, intraobserver variability was r= .96 • no statistically significant correlation found between partial volume-corrected SUV values of invasive breast cancer and tumor size, blood glucose level, tumor-cell proliferation and histopathologic grading • higher SUV values for ER-negative tumors compared with ER-positive tumors, but not statistically significant • no statistically significant correlation between SUV values of breast cancer and PR status **Authors' comments** • partial volume effects in tumors < 1 cm in size limits detection by PET • in situ carcinoma showed an increase in FDG uptake lower than that of invasive cancer, but further study in larger populations are needed • further studies are needed to determine the prognostic value of FDG uptake in breast cancer

Study	Patients/Methods	Results/Comments
Scheidhauer, et al., 1996 *University of Cologne and Max-Planck Institut for Neurological Research, Cologne, Germany)*	**Purpose** • to assess the diagnostic accuracy of qualitative PET scans • to demonstrate imaging results in a manner acceptable to referring clinicians • to minimize scanning time and patient discomfort **Cases** 30 patients with suspicion of breast cancer based on clinical exam or mammography/ ultrasonography (23 malignant, 7 benign) • 18 of whom also had ipsilateral axillary node exploration; no axillary exploration done in 5 patients with locally advanced disease who received neoadjuvant chemotherapy before surgery • 8 patients with distant metastases at the time of diagnosis **Methods** • patients fasted at least 12 hours before PET • PET scans performed on patients in the supine position • attenuation-corrected emission images and transmission images available to investigator; data displayed on computer screens as 3 orthogonal images and with interactive choice of slice localization by investigator; data also viewed by Multi Purpose Matching (MPM) software • qualitative PET images evaluated by two readers, blinded to all information other than that the patient was scheduled for breast surgery • surgery performed between 1 and 5 days after PET scans • all breast lesions biopsied; additional foci of increased FDG uptake not corresponding to area of suspicion on other imaging also biopsied • PET, clinical exam, and mammography compared to histology **Limitations of study design** • number of cases and internal controls not equivalent (high index of suspicion for malignancy) • PET results and determination of disease status not independent • study design methods (i.e., blinding) for mammography and palpation not described	**Detecting primary tumors (23 malignant, 7 benign)** PET: Se=91%; Sp=86% Mammography: Se=86% Clinical exam: Se=74%; Sp=71% **Detecting axillary lymph node involvement (9 malignant, 9 benign)** PET: Se=100%; Sp=89% **Detecting distant metastases (8 positive, 15 negative; 8 total patients in the PET field of view)** PET: Se=100%; Sp=100% **Authors' comments** • a more time-effective qualitative approach does not decrease the accuracy of breast cancer detection • MPM software did not enhance accuracy when compared with standard displays of orthogonal slices (no data reported) • combination functional scan with anatomical scan showed higher acceptance by referring physician and improved intraoperative orientation, thereby facilitating the identification and correct removal of the area of interest (no data reported) • supine position during imaging reflects position during surgery • because of selection criteria, accuracy data cannot be used to judge the accuracy of these techniques, but FDG PET may yield additional information on tumor biology • authors suggest PET is comparable to other imaging techniques with respect to imaging time and scanning discomfort • authors suggest roles for PET in patients with inconclusive findings prior to biopsy, for preoperative TMN staging in patients with highly suspicious breast findings to aid therapy planning, and to replace palpation and conventional imaging tools when they are technically not feasible (eg. in patients with silicone impants)

Abbreviations:
DUR, dose uptake ratio
Se, sensitivity
Sp, specificity
TNT, tumor to normal uptake ratio
SUV, standardized uptake value
ROI, region of interest

* indicates calculated by MDRC TA Program from data supplied in published article

VI. REFERENCES Background and studies meeting evidence-based medicine criteria for evaluations of diagnostic tests

Adler DD, Wahl RL. New methods for imaging the breast: techniques, findings, and potential. *American Journal of Roentgenology*. 1995;164:19-30.

Adler LP, Crowe JP, Al-Kaisi NK,Sunshine JL. Evaluation of breast masses and axillary lymph nodes with [f-18] 2-deoxy-2-fluoro-D-glucose PET. *Radiology*. 1993;187:743-50.

Albertini JJ, Lyman GH, Cox C, Yeatman T, Balducci L, Ku N, et al. Lymphatic mapping and sentinel node biopsy in the patient with breast cancer. *JAMA*. 1996;276(22):1818-22.

American Cancer Society. *Cancer Facts & Figures-1996*. New York: National Media Office-ACS, 1996.

Avril N, Dose J, Jänicke F, Ziegler S, Römer W, Weber W, et al. Assessment of axillary lymph node involvement in breast cancer patients with positron emission tomography using radiolabeled 2-(fluorine-18)-fluoro-2-deoxy-D-glucose. *Journal of the National Cancer Institute.* 1996;88(17):1204-9.

Avril N, Dose J, Jänicke F, Bense S, Ziegler S, Laubenbacher C, et al. Metabolic characterization of breast tumors with positron emission tomography using F-18 fluorodeoxyglucose. *Journal of Clinical Oncology*. 1996;14:1848-57.

Blamey RW, Wilson ARM, Patnick J, Dixon JM. Screening for breast cancer. *British Medical Journal*. 1994;309:1076-9.

Department of Veterans Affairs, National Center for Veteran Analysis and Statistics, Assistant Secretary for Policy and Planning: *Nation Survey of Veterans (NSV9503)*. April 1995. Depot Stock No. P92493.

Feussner JR, Hynes DM. Breast cancer among women veterans -- pilot/feasibility study. SDR 92-006, Abstract. In: HSR&D Special Projects Office: *Projects Receiving Funding in Fiscal Year 1995.*

Harms SE, Flamig DP, Evans WP, Harries SA, Bown S. MR imaging of the breast: current status and future potential. *American Journal of Roentgenology*. 1994;163:1039-47.

Holle LH, Trampert L, Lung-Kurt S, Villena-Heinsen CE, Püschel W, Schmidt S, et al. Investigations of breast tumors with fluorine-18-fluorodeoxyglucose and SPECT. *The Journal of Nuclear Imaging.* 1996;37:615-22.

Institute of Medicine. *Strategies for managing the breast cancer research program: a report to the U.S. army medical research and development unit.* Washington, D.C.: National Academy Press; 1993.

Kelsey JL, Horn-Ross PL. Breast cancer: magnitude of the problem and descriptive epidemiology. *Epidemiologic Reviews.* 1993;15:7-16.

Liff JM, Sung JFC, Chow WH, Greenberg RS, Flanders WD. Does increased detection account for the rising incidence of breast cancer? *American Journal of Public Health.* 1991;81:462-5.

Minn H, Soini I. [^{18}F]Fluorodeoxyglucose scintigraphy in diagnosis and follow up of treatment in advanced breast cancer. *European Journal of Nuclear Medicine.* 1989;15:61-6.

Morrison AS. Screening for cancer of the breast. *Epidemiologic Reviews.* 1993;15:244-55.

Nieweg OE, Kim EE, Wong WH, Broussard WF, Singletary SE, Hortobagyi GN, et al. Positron emission tomography with fluorine-18-deoxyglucose in the detection and staging of breast cancer. *Cancer.* 1993;71:3920-5.

Parker SH, Burbank F, Jackman RJ, Aucreman CJ, Cardenosa G, Cink TM, et al. Percutaneous large-core breast biopsy: a multi-institutional study. *Radiology.* 1994;193:359-64.

Scheidhauer K, Scharl A, Peitrzyk U, Wagner R, Göhring UJ, Schomäcker K, et al. Qualitative [^{18}F]FDG positron emission tomography in primary breast cancer: clinical relevance and practicability. *European Journal of Nuclear Medicine.* 1996;23(6):618-23.

Tubiana M, Holland R, Kopans DB, Kurtz JM, Petit JY, Rilke F, et al. Commission of the european communities “europe against cancer” program. european school of oncology advisory report: management of non-palpable and small lesions found in mass breast screening. *European Journal of Cancer.* 1994;30A;538-47.

White E, Urban N,Taylor V. Mammography utilization, public health impact, and cost effectiveness in the United States. *American Review of Public Health.* 1993;14:605-33.

VII.REFERENCES: Technical efficacy studies

Bassa P, Kim EE, Inoue T, Wong FCL, Korkmaz M, Yang DJ, et al. Evaluation of preoperative chemotherapy using PET with fluorine-18-fluorodeoxyglucose in breast cancer. *The Journal of Nuclear Medicine.* 1996;37:931-8.

Bruce DM, Evans NTS, Heys SD, Needham G, BenYounes H, Mikecz P, et al. Positron emission tomography: 2-deoxy-2-[^{18}F]-fluoro-D-glucose uptake in locally advanced breast cancers. *European Journal of Surgical Oncology.* 1995; 21:280-3.

Jansson T, Westlin JE, Ahlström H, Lilja A, Långström B,Bergh J. Positron emission tomography studies in patients with locally advanced and/or metastatic breast cancer: a method for early therapy evaluation? *Journal of Clinical Oncology.* 1995;13:1470-7.

Mortimer JE, Dehdashti F, Siegel BA, Katzenellenbogen JA, Fracasso P, Welch MJ. Positron emission tomography with 2-[^{18}F]fluoro-2-deoxy-D-glucose and 16a-[^{18}F]fluoro-17ß-estradiol in breast cancer: correlation with estrogen receptor status and response to systemic therapy. *Clinical Cancer Research.* 1996;2:933-9.

Pietrzyk U, Scheidhauer K, Scharl A, Schuster A, Schicha H. Presurgical visualization of primary breast carcinoma with pet emission and transmission imaging. *The Journal of Nuclear Medicine.* 1995;36(10):1882-4.

Wahl RL, Cody RL, Hutchins GD, Mudgett EE. Primary and metastatic breast carcinoma: initial clinical evaluation with PET with the radiolabeled glucose analogue 2-[f-18]-fluoro-2-deoxy-D-glucose. *Radiology.* 1991;179:765-70.

Wahl RL, Zasadny K, Helvic M, Hutchins GD, Weber B, Cody R. Metabolic monitoring of breast cancer chemohormotherapy using positron emission tomography: initial evaluation. *Journal of Clinical Oncology.* 1993;11(11):2101-11.

Zasadny KR, Wahl RL. Standardized uptake values of normal tissues at PET with 2-[fluorine-18]-fluoro-2-deoxy-D-glucose: variations with body weight and a method for correction. *Radiology*. 1993;189:847-50.

VIII. REFERENCES: Excluded studies

Exclusion criteria included:

- number of cases < 12
- duplicated or superseded by subsequent or concurrent study from the same institution
- radiopharmaceutical other than FDG
- gamma camera rather than PET
- tumors other than squamous cell carcinomas
- insufficient information to judge comparability of case and control groups, details of imaging protocol, whether visual or quantitative analysis of PET data used, or type of PET data analysis used
- abstract, not peer reviewed

Bombardieri E, Crippa F, Maffioli L, Chiti A, Castellani MR, Greco M, et al. Axillary lymph node metastases detection with nuclear medicine approaches in patients with newly diagnosed breast cancer: can positron emission tomography (PET) with ^{18}F-FDG be considered as the best method? *International Journal of Oncology*. 1996;8:693-9.

Dehdashti F, Mortimer JE, Siegel BA, Griffeth LK, Bonasera TJ, Fusselman MJ, et al. Positron tomographic assessment of estrogen receptors in breast cancer: comparison with FDG-PET and in vitro receptor assays. *The Journal of Nuclear Medicine*. 1995;36(10):1766-74.

Hoh CK, Hawkins RA, Glaspy JA, Dahlbom M, Tse NY, Hoffman EJ, et al. Cancer detection with whole-body PET using 2-[^{18}f]fluoro-2-deoxy-D-glucose. *Journal of Computer Assisted Tomography*. 1993;17(4):582-9.

McFarlane DJ, Cotton L, Ackermann RJ, Minn H, Ficaro EP, Shreve PD, et al. Triple-head SPECT with 2-[fluorine-18]fluoro-2-deoxy-D-glucose (FDG): initial evaluation in oncology and comparison with FDG PET. *Radiology*. 1995;194:425-9.

Mintun MA, Welch MJ, Siegel BA, Mathias CJ, Brodack JW, McGuire AH, et al. Breast cancer: PET imaging of estrogen receptors. *Radiology*. 1988;169:45-8.

Nieweg OE, Wong WH, Singletary SE, Hortobagyi GN, Kim EE. Positron emission tomography of glucose metabolism in breast cancer: potential for tumor detection, staging, and evaluation of chemotherapy. *Annals New York Academy of Sciences*. 1993;689:423-8.

Tse N, Hoh CK, Hawking RA, Zinner MJ, Dahlbom M, Choi Y, et al. The application of positron emission tomographic imaging with fluorodeoxyglucose to the evaluation of breast disease. *Annals of Surgery*. 1992;216:27-34.

Wahl RL, Helvie MA, Chang AE, Andersson I. Detection of breast cancer in women after augmentation mammoplasty using fluorine-18-fluorodeoxyglucose-PET. *The Journal of Nuclear Medicine*. 1994;35:872-5.

Wahl RL, Kaminski MS, Ethier SP, Hutchins GD. The potential of 2-deoxy-2[^{18}f]fluoro-D-glucose (FDG) for the detection of tumor involvement in lymph nodes. *The Journal of Nuclear Medicine*. 1990;31:1831-5.

Wahl RL, Cody, August D. Initial evaluation of FDG-PET for the staging of the axilla in newly diagnosed breast carcinoma patients. *The Journal of Nuclear Medicine.* 1991;32:981.

Wahl RL, Cody R, Hutchins GD, Mudgett EE. Positron emission tomography scanning of primary and metastatic breast cancer with the radiolabeled glucose analogue 2-deoxy-2-[^{18}F]fluoro-D-glucose. *New England Journal of Medicine*. 1991;324:200.

Appendix 6

Systematic Review: PET as a Diagnostic Test in Lung Cancer

Author: Elizabeth Adams, R.R.T., M.P.H., *Management & Program Analyst, MDRC Technology Assessment Program*

Appendix 6

Systematic Review: PET as a Diagnostic Test in Lung Cancer

The final literature database searches for the systematic reviews were performed on September 10, 1996; the assessment represents peer-reviewed literature published and indexed as of that date.

This Appendix to the PET assessment presents the results of the systematic review of PET in lung cancer. A general rationale for the use of PET in oncology is supplied by Hawkins, et al., (1994) and Hoh, et al., (1994):

- many forms of cancer characteristically perturb tissue biochemical and physiological processes and PET imaging can be expected to detect the resulting abnormalities;
- reliance on tumor histology and anatomy limits the oncologist's tools for selecting optimal treatment;
- the ability to monitor metabolic responses to treatment could allow the early re-direction of therapy in patients who fail to respond to the first attempt at radiation or chemotherapy.

These and other authors (e.g., Price and Jones, 1995) report that PET studies in cancer are emerging as a major focus of the technology, both in basic research and in clinical investigations. Information gathered by the MDRC Technology Assessment Program from VA PET facilities corroborates that perception (see *Appendix 9: Experience With PET in VHA*).

Fluorine-18-fluorodeoxyglucose (FDG) is the most commonly employed radiopharmaceutical in PET cancer studies. Many neoplasms have high glycolytic rates, resulting in intracellularly trapped phosphorylated FDG that can be imaged with PET. Hawkins, et al., (1994), note that tumor-specific biochemical characteristics of glucose transport and phosphorylation may affect quantitative estimates of tumor glucose metabolism with FDG PET, and that investigations are under way to define these characteristics. However, these uncertainties may be of less concern with qualitative or semiquantitative FDG PET cancer studies because the primary intent of such studies is to detect and map tumor foci, not to rigorously quantify tumor glycolytic rates.

In some instances, PET imaging techniques have been modified to meet the needs of cancer diagnosis. Most PET systems allow axial fields of view (the length of the body encompassed by a series of cross sectional images) of approximately 10 cm. Cancer is frequently distributed beyond this field of view, and whole body image acquisition procedures have been developed (Hoh, et al., 1993). Since it is impractical to apply standard transmission scanning attenuation correction methods to these procedures, whole body PET imaging is primarily useful as a qualitative indicator of disease distribution.

Nieweg (1994) and Price and Jones (1995) define a number of potential applications for PET in oncology. These include:

- tumor detection (although PET images offer insufficient structural detail and should not be used to visualize anatomy; registration techniques to combine PET and anatomic imaging into a single image are under development to circumvent this limitation);

- staging (particularly using whole-body imaging methods) although there is a lower limit to the size of metastases that can be detected by PET;

- detection of local recurrence of disease, since anatomically-based imaging is often limited by the effects of treatment;

- prediction of tumor response to chemotherapy;

- treatment monitoring.

I. BACKGROUND

A. General sources

The information in this section, unless otherwise noted, is based on Minna (1994). Additional sources are referenced in the text.

B. Description

Bronchogenic carcinoma, classified as either small cell or non-small cell, comprises 95% of all primary lung cancers. Three-fourths of all bronchogenic carcinomas are of the non-small cell varieties and include cell types which, when localized, have the potential for cure with surgical resection. They include adenocarcinoma (including bronchiolalveolar), squamous (or epidermoid) cell carcinoma, and large cell (including large cell anaplastic) carcinoma. This report will not address small cell lung carcinomas, because they occur less frequently and are staged and treated differently than non-small cell types.

C. Epidemiology

Bronchogenic carcinoma is the leading cause of cancer death in the United States. In 1996 it is estimated that there will be 177,000 new cases of primary lung carcinoma and 158,700 deaths from lung cancer (American Cancer Society, 1996). Within the Veterans Health Administration, malignant neoplasms of the bronchus and lung accounted for a total of 14,749 patients discharged (1.75% of all patients discharged within the system) with an average length of stay of 18.0 days in fiscal year 1994 (Annual Report of the Secretary of Veterans Affairs, 1994).

The overall incidence is increasing, causing the age-adjusted lung cancer death rate to double every 15 years. The major risk factor for all lung cancers is smoking. The contribution of second-hand smoke is controversial, but is estimated to be responsible for 15% to 20% of lung cancers in non-smokers (Filderman, 1994). There is a dose-response correlation between lung cancer death rate and the total amount of cigarettes smoked. Likewise, cessation of smoking decreases the risk of developing lung cancer, although the risk may never return to normal levels. Additional risk factors may include exposure to: asbestos, chromium, nickel, mustard gas, vinyl chloride, arsenic, isopropyl oil, hydrocarbons, radon, and chloromethyl ether, and ionizing radiation from occupational, medical, and environmental sources (Filderman, 1994). Most, and perhaps all, of these materials are additive or synergistic with cigarette smoke in the development of lung cancer.

Evidence suggesting a genetic predisposition to lung cancer has been reported, but the underlying mechanisms have not yet been identified (Samet, 1993). Because lung cancer rarely occurs in the absence of tobacco exposure, host characteristics may be expressed only in the presence of an environmental insult, such as tobacco.

Dietary factors may contribute independently to the incidence of lung cancer in some populations. While a direct association between increased consumption of dietary cholesterol and animal fat and increased lung cancer risk has been reported, current evidence supports neither benefit nor harm from the use of supplemental beta carotene (Kabat, 1993; Zagonel, 1994; Hennekens, 1996).

D. Diagnosis

In unscreened and asymptomatic patients, 5-15% of non-small cell lung cancers are detected on a routine chest radiograph usually ordered for other reasons. However, the vast majority of patients are symptomatic, indicating advanced disease at clinical presentation.

The clinical manifestations of lung cancer vary with cell type and extent of tumor spread (stage), and may be confused with paraneoplastic syndromes (a group of symptoms resulting from nonmetastatic complications that may mimic metastatic disease). Signs and symptoms resulting from local tumor spread include pain and discomfort resulting from tumor involvement of adjacent thoracic structures such as the heart, esophagus, trachea and chest wall. More severe symptoms may include respiratory insufficiency and impaired oxygenation.

Initial diagnosis is based on a complete history, physical examination, and chest radiography (planar x-ray). If results of the chest film increase the likelihood of cancer, resulting in a high post-test probability of disease, other tests are needed to determine stage, cell type, and subsequent treatment. Patient tolerance, tumor accessibility, and the risks and costs associated with each available test will determine the best method(s) (i.e., degree of invasiveness needed) for obtaining tissue specimens to optimize diagnostic certainty. Any one or combination of the following methods may be used to obtain tissue specimens: sputum sampling, diagnostic bronchoscopy, percutaneous transthoracic biopsy, mediastinoscopy, thoracoscopy, and thoracotomy (Kaplan, 1991). These endoscopic procedures and thoracotomy are also used to further visualize extent of disease.

E. Staging, treatment, and survival

Lung cancer staging assesses the extent of local and distant disease and involves two parts: 1) anatomic staging, and 2) physiologic staging or the ability of the patient to tolerate specific therapeutic interventions (performance status). Illustrated below is the TNM International Stage System (ISS) developed by the American Joint Committee on End Results Reporting used to describe the extent of primary tumor involvement (T stage), lymph node involvement (N stage) and distant metastasis (M stage), and to reflect prognosis and survival among homogenous patient groups with non-small cell lung carcinomas (Mountain, 1993).

Table 1 Lung Cancer TNM Staging System

<table>
<tr><th></th><th>N_0</th><th>N_1</th><th>N_2</th><th>N_3</th><th></th></tr>
<tr><td>T_0</td><td>carcinoma in situ</td><td colspan="3"></td><td>Roman numerals represent stages</td></tr>
<tr><td>T_1</td><td rowspan="2">I</td><td rowspan="2">II</td><td rowspan="2"></td><td rowspan="3"></td><td>Occult stage=$T_xN_0M_0$</td></tr>
<tr><td>T_2</td><td>Occult stage through stage IIIb are without distant metastases (M_0)</td></tr>
<tr><td>T_3</td><td colspan="3">I I I a</td><td>All M_1 tumors (with distant metastases) are stage IV</td></tr>
<tr><td>T_4</td><td colspan="4">I I I b</td><td></td></tr>
</table>

subscripted numbers=degree of involvement; 0=least to 4=most

Source: Mountain, 1993

The data in the following table were provided by NCI through its on-line Physician Data Query (PDQ) system to present staging, treatment, and survival data for patients with non-small cell lung cancer.

Table 2 Lung Cancer Staging, Treatment, and Survival

Primary Site	***Staging***	***Standard Therapy***	***5-Year Survival (Other Therapy-Specific Survival Data, Where Indicated)***
Lung Parenchyma	Occult	Surgery	70-80%, overall for occult
	0= *in situ*	surgery with curative intent photodynamic therapy	not available
	I	surgery with curative intent radiotherapy with curative intent, depending on T-stage neoadjuvant therapy *5-10% of patients may develop second lung cancers within 5 years*	50% overall for stage 1 10-60%
	II	surgery with curative intent radiotherapy with curative intent, depending on T-stage surgery and/or radiotherapy and/or chemotherapy	30% overall for stage 11 10-60%
	IIIa	surgery surgery and radiotherapy radiotherapy surgery and/or radiotherapy plus chemotherapy	10-30% overall for stage 111a
	IIIb	radiotherapy with curative intent chemotherapy plus radiotherapy chemotherapy plus radiotherapy followed by surgery chemotherapy alone radiotherapy followed by surgery	< 5% overall for stage 111b
	IV	radiotherapy with palliative intent chemotherapy, depending on performance status chemo- and radiation therapy	< 2% overall for stage IV
	Recurrence	radiotherapy chemotherapy both	not available

Source: NCI, 1995

At the time of diagnosis approximately 55% of patients with lung cancer will have stage IV disease; 25% will have either stage II, IIIa or IIIb disease; and 20% will have local stage I disease. Surgical resection is the treatment of choice for patients with operable lung cancer (occult carcinoma through stage IIIa). For inoperable stage IIIa, IIIb and IV cancer, radiotherapy is the preferred option for palliation in patients with poor performance status or in patients who refuse multimodality regimines, but it results in cure for only a small minority of patients. In stage IIIb and IV disease chemotherapy offers modest improvements in patients with a good performance status, although overall survival is poor. The effect of chemotherapy on survival in patients with poor performance status is unknown (Souquet, 1993). Multimodality treatment in more advanced stages of disease to improve survival is being studied.

Recent advances in endoscopic surgical equipment, surgical techniques, and neoadjuvant (preoperative) chemotherapy with and without radiotherapy aid in converting some patients from unresectable to resectable status. Unfortunately, advances in staging methods and surgical treatment of lung cancer have had little impact on overall mortality rates, although more accurate staging, particularly of the mediastinum, has significantly reduced the incidence of unbeneficial exploratory thoracotomy (Pearson, 1993). Two ongoing randomized trials, one evaluating extensive versus limited resection and the other evaluating the effect of surgery in more extensive disease, may provide insight into the effectiveness of surgery in the treatment of localized non-small cell lung cancer (Lederle, 1994).

Prognostic factors of patients with non-small cell lung cancer have been identified as those predictive of treatment response and those predictive of survival (Shepherd, 1994). As in all malignancies, the primary predictor of response to treatment is the stage of disease at diagnosis; others include performance status and chemotherapeutic regimen (eg., single agent versus combination therapy). The most significant factors predictive of survival are stage at diagnosis and performance status followed by gender, history of pretreatment weight loss, and/or elevated blood lactate dehydrogenase levels. Since the majority of patients present with advanced stage disease, there is a need to identify (or “screen”) individuals, particularly those at high risk, at a point early in tumor development with the hope that the tumor would still be amenable to curative surgical resection.

Lung cancer screening is usually performed using serial chest radiography (x-ray) and sputum sampling. However, studies using these screening tools have not demonstrated a clear survival benefit because of their low sensitivity. Moreover, results were confounded by the effects of either the surgical intervention, lead time bias (the interval between the diagnosis of a disease at screening and when it would have been detected due to development of symptoms), or length-time bias (overrepresentation among screen-detected cases of slower growing tumors which have a more favorable prognosis). Accordingly, there is a need to develop other technologies to overcome these limitations.

The prognostic role of molecular diagnostics as an adjunct to lung cancer screening tools represents a significant portion of research activity to date. Tumor biomarkers may provide insight into the natural history of occult or pre-cancerous tumor development and their corresponding treatment. Advancements in fiberoptic technology enhanced with laser (eg., fluorescence bronchoscopy) are being developed to locate pre-cancerous cells in the airways, which would enable the patient to undergo curative resection much earlier in the course of the disease. Methods for detecting genetic alterations, including oncogene anomalies and deoxyribonucleic acid (DNA) mutations in sputum and bronchoscopic specimens, in patients with lung cancer are areas of active investigation.

F. Potential roles for PET

Currently, CT is the preferred diagnostic imaging test and is used at several points in the initial work up and treatment of a patient with lung cancer. Its roles include staging, evaluating treatment response, and differentiating recurrent tumor from fibrosis after treatment. However, its limitations are well known. CT provides morphologic (typically size), not histologic, detail of the disease site. Therefore, biopsy confirmation of the primary site and metastases is required to determine the most appropriate treatment.

The wide range of reported accuracies makes the contribution of CT to lung cancer staging difficult to quantify. The following factors are likely to influence the reported characteristics of CT: differences in disease prevalence among study populations; cell type; scanning techniques; definition of the boundary between adjacent node structures; criteria for lymph node enlargement; data analysis (by patient or nodal station); and extent of node sampling performed either pre- or peri-operatively (Quint, 1995 and Seely, 1993). The contribution of interobserver variability in image interpretation has also been identified (Guyatt, 1995 and Webb, 1993).

Detection of cancer in mediastinal lymph nodes is particularly problematic. CT of the mediastinum may demonstrate the presence of enlarged but benign lymph nodes, and may often appear normal in the presence of micrometastases. A meta-analysis assessing the use of CT in staging lung cancer found CT to be 80% accurate in evaluating mediastinal lymph nodes, and advances in CT staging techniques in recent years have had little measurable affect on accuracy (Dales, 1990). While predictive values are generally more helpful in the clinical management of these patients, the range of prevalence of malignancy in these studies precluded the use of predictive values in their analysis. Using data based on sensitivity, specificity, and accuracy the authors recommended that an indicator other than lymph node size be used to determine lymph node pathology.

Determination of distant metastases (M stage) may require multiple scans, which can be very resource-intensive. Brain and abdominal CT and radionuclide bone scanning are most often employed as part of the work up for metastatic disease. Standard practice supports the use of scanning those organs in patients demonstrating symptoms of metastatic disease, but routine use in patients with an otherwise unremarkable clinical exam remains controversial. A systematic review conducted by Hillers, et al. (1994) confirmed the controversial nature of the metastatic work up in these patients, although the prevalence of unsuspected brain metastases in patients with lung cancer may provide the rationale among some clinicians for routine brain CT scanning.

A meta-analysis conducted by Silvestri, et al., (1995) concluded that the negative predictive value of an unremarkable clinical evaluation when compared with brain CT, abdominal CT, or radionuclide bone scanning consistently exceeded 90%. The negative predictive value exceeded 97% when an expanded pre-defined set of criteria was added to the routine clinical evaluation of metastatic disease to the brain or abdomen. A decision analysis conducted by this same group further supported not using brain CT scans routinely in the presence of normal findings on clinical examination (Colice, 1995). Both studies reemphasized the importance of a standardized physical examination, history, and basic lab tests in staging these patients.

Use of other diagnostic imaging technologies in the staging of lung cancer is circumscribed largely because of technical limitations, availability, and cost. Whereas MRI has not demonstrated additional benefit over CT or in combination with CT in staging, it may help in delineating vascular structures within the hila and mediastinum and in detecting aortopulmonary and subcarinal lymphadenopathy. Excellent visualization provided by coronal, sagittal, and oblique views of the chest does offer MRI several advantages over CT in staging Pancoast tumors and imaging tumor invasion of both diaphragmatic surfaces and the chest wall. Ultrasonography may also have a limited role in detecting tumors invading the chest wall.

Advances in nuclear medicine imaging have focused on qualitative and quantitative physiologic, rather than morphologic, determination of disease status with the intent to improve the accuracy of diagnosis. So far, the results of nuclear medicine studies using Gallium-67, Thallium-201, and Technetium-99m-sestaMIBI (Chiti, 1996) to stage lung cancer have not demonstrated marginal benefit over CT. The use of immunoscintigraphy (radiolabelled monoclonal antibodies) in early detection of lung cancer is being applied, but validation with larger trials is needed.

Potential roles for whole body FDG PET in lung cancer have been noted in the literature (Nieweg, 1994 and Gupta, 1993). PET has been evaluated in the detection of the primary tumor, staging, and distinguishing recurrent disease from scar tissue. Interpretation of these PET studies is accomplished by visual inspection and by various quantitative or semiquantitative analyses used to characterize disease status.

Results of FDG PET imaging in patients with lung cancer were first published by Nolop, et al., (1987) from the Royal Postgraduate Medical School, London, England. This study demonstrated that the quantitative assessment of glucose utilization in pulmonary neoplasms is feasible and may have important therapeutic implications. Rege, et al., (1993) from UCLA School of Medicine were the first to demonstrate the feasibility and potential utility of the whole body PET method to image primary and metastatic chest tumors.

Additional roles for PET have been suggested. They include:

- analyzing tumor biology;
- predicting tumor response by measuring uptake of chemotherapeutic agents;
- quantitatively monitoring tumor response to therapy.

The MDRC TA Program was unable to identify any studies which evaluated these roles in lung cancer and which met the screening criteria for this assessment.

II. RESULTS

Thirty-seven articles were selected from the MEDLINE and other database searches and from the bibliographies of initially retrieved articles as meeting the screening criteria. After review, 21 (57%) were found to meet the criteria for assignment to the following levels of the diagnostic efficacy hierarchy (Fryback and Thornbury, 1991; *Appendix 2: Assessing Diagnostic Technologie*s): 9 met the definition of technical efficacy (see Reference List; full data abstraction tables for Technical Efficacy studies are on file with the MDRC Technology Assessment Program); 11 met most or all of the criteria for studies of diagnostic efficacy; 1 met the criteria for studies of diagnostic efficacy and attempted to address diagnostic thinking efficacy hypothetically.

Tables 4 and 5 abstracted data from studies of diagnostic accuracy of PET for certain lung cancer applications. Table 3 summarizes cross-study findings on PET and alternative technologies. The MDRC Technology Assessment Program was unable to locate any studies using PET in lung cancer at the patient outcome or societal levels.

Gambhir, et al., (1996) presented a decision analysis for the cost-effectiveness of FDG-PET in the staging and management of non-small cell lung cancer. This study was not included in the tables because of the preliminary nature of the assumptions upon which the analysis was based, which may affect the stability of the conclusions. Sensitivity and specificity values used in the analysis were derived from 3 small preliminary studies (two abstracts, one peer-reviewed) comprising 96 total patients analyzed by nodal station and mediastinal side. Cost data were based on billable costs (charges) that may not adequately reflect true costs, and that may not have been sufficiently comprehensive with respect to inclusion of other costs related to the work-up and to patients with unresectable metastases. The results from this analysis may not be valid and should be viewed with caution.

All currently available data on the use of PET in diagnosing lung cancer are derived from case series studies. These studies provide Level V (i.e., the weakest) evidence of any association between the use of a technology and improved patient outcomes. All studies included patients with a high index of suspicion for lung cancer and used internal controls (patients with benign masses). Accordingly, no predictive values were reported, with the exception of Knight, et al., (1996), who had an equivalent array of subjects with which to calculate predictive values for a small subset of patients.

PET was evaluated at various points in the test sequence in the diagnosis of lung cancer either as an addition to or as a substitute for CT. Results for the use of PET in detecting unknown primary disease, nodal metastases, and recurrent disease are presented below. Anecdotal data on the use of PET in detecting distant metastases were presented by Valk, et al., (1995) in Table 4, but were not included in Table 3 because of the small number of patients.

A. Detecting unknown primary disease

Table 4 lists six preliminary studies using PET in the diagnosis of primary lung cancer, of which two (Wahl, 1994; Sazon, 1996) assessed PET and CT independently. The remaining four studies evaluated the complementary role of PET with CT in the diagnostic process. One study by Valk, et al., (1995) did not report operating characteristics for either PET or CT in detecting primary disease and was not included. Only two studies (Wahl, 1994 and Knight, 1996) presented CT data for comparison; limitations in study design and small study size call for cautious interpretation of these results.

In most studies the extent to which the PET results influenced determination of disease (eg., to proceed to thoracotomy versus follow-up) could not be ascertained. Of interest is a finding by Kubota, et al., (1990), who noted that the differences in operating characteristics between the two PET models used in their study may have affected the generalizability of reported results.

Table 5 lists one study (Slosman, 1993) that assessed the impact of PET on diagnostic thinking efficacy using Bayesian analysis in patients seen at a satellite center for a diagnostic work up of lung cancer. Using sensitivity figures derived from their own study population, but specificity from Kubota, et al., (1990), and varying the prevalence of disease, the authors hypothesized the impact of PET on diagnostic certainty. While they illustrated that PET would have the greatest impact in a population with a low prevalence of disease, and in whom a positive test could lead to more aggressive therapy than one would otherwise plan, further study in larger populations is needed to define the role of PET in the diagnostic work-up.

B. Detecting hilar and mediastinal metastases

Several studies evaluated PET at various points in the test sequence for staging mediastinal involvement in lung cancer. Most studies evaluated PET qualitatively, while Scott, et al., (1996) evaluated PET quantitatively using a cut-off value based on unpublished data.

Variations in study design influenced the range of reported results and contributed to the degree of bias found in these studies. These relatively small studies comprised a narrow range of patients restricted to biopsy-proven cases with a high index of suspicion for metastases. In three studies (Patz, 1995; Sazon, 1996; Scott, 1996) the choice of patients included for mediastinal staging was influenced by the PET results.

CT and PET are limited in their ability to detect micrometastases and require histologic confirmation of disease. In most, and perhaps all, of these studies there was a strong association between the test result and choice of biopsy site. Variations in biopsy sampling procedures and the thoroughness of sampling, often left to the discretion of the surgeon, occurred across all studies and typically was not reported with sufficient detail to be reproducible. While standard practice supports the use of imaging by surgeons either pre- or peri-operatively, knowledge of imaging test results may favor nodes that appear suspicious on imaging, resulting in biased test characteristics.

Differences in methods of data analysis reported by patient, by mediastinal side, and by node were presented and may contribute to the range of reported results. Scott and associates (1996) presented data by both patient and node. Operating characteristics from all of these studies should be interpreted cautiously, as the degree of bias is significant.

C. Detecting recurrent disease

Two studies (Patz, 1994 and Inoue, 1995) used PET with CT to distinguish recurrent or residual cancer from fibrosis. Both studies incorporated a semiquantitative methodology to characterize comparison groups. The preliminary nature of these studies must be stressed because of variations in cut-off points, one which was determined retrospectively, and low number of study subjects.

III. SUMMARY

Preliminary studies of the potential roles of FDG PET in diagnosing lung cancer using visual inspection, semiquantitative analysis, and Bayesian analysis were presented. None of these studies demonstrated the incremental value of PET in the sequence of tests used to diagnose and stage lung cancer or to distinguish local cancer recurrence from fibrosis. All were relatively small case series using internal controls with a disproportionately high number of malignant cases, and may not provide reliable estimates of accuracy. The MDRC Technology Assessment Program was unable to locate any studies in lung cancer that evaluated the incremental value of PET information on treatment planning or patient outcome. Gambhir and associates (1996) presented a decision analysis to assess the cost-effectiveness of PET in the work-up of non-small cell lung cancer, but the underlying assumptions used in the study may not be valid.

None of the studies met strict evidence-based medicine criteria for blinding, and there is a strong likelihood that the test results may have influenced the determination of disease status. However, all studies presented information on the comprehensiveness of blinding of the test interpreters to the diagnostic gold standard (i.e., biopsy confirmation). One small, hypothetical diagnostic thinking efficacy study (Slosman, 1994) was included, but further validation with larger study populations is needed. Table 3 summarizes the findings from studies assessing the diagnostic accuracy of PET. These findings received low methodologic quality grades because of small study size, retrospective design, and a significant degree of bias.

IV. DISCUSSION

The attempt by some of these studies to characterize comparison groups quantitatively warrants further discussion. Consideration of the cut-off point used in quantitative analysis will depend on the consequences of limiting false negative results and of accepting false positive results. This value will also be influenced by the pre-test probability of disease of the study population and the heterogeneity of both normal and cancerous lung tissue. In a recently published case control study Miyauchi, et al., (1996) demonstrated the effect of regional variations of FDG uptake within normal lungs on the range of reported results, particularly with respect to small lung nodules found in lower lung fields. Variations in normalization procedures used in semi-quantitative analyses may further influence the choice of cut-off (Schomburg, 1996). Determination of the clinical efficacy of PET using quantitative methods requires first standardizing the technique, defining the optimal cut-off point from a much larger and broader study population, and subsequently applying it to studies designed to determine diagnostic accuracy.

One study attempted to evaluate the impact of PET on diagnostic certainty using Bayesian analysis (Slosman, et al., 1993). Although this study has methodologic shortcomings (i.e., an insufficient number of controls with which to determine specificity), it does illustrate the importance of defining the baseline prevalence of disease within the study population of interest when evaluating operating characteristics of a test.

As modifications of more widely available and less costly alternatives are refined and evaluated, their contribution to the medical workup may affect the utility of PET as a diagnostic tool for lung cancer. Over the last two decades, there have been considerable developments in medical imaging, many of which use variations of the techniques used in CT to improve diagnostic accuracy. An increased understanding of the physics underlying imaging, enhanced computer capabilities, and applied mathematical-reconstruction techniques have made a significant contribution to improved image generation and analysis (Greenes and Brinkley, 1990). Most, if not all, of the techniques described below are in the preliminary stage of evaluation, and their contribution to the diagnostic process has not been quantified.

New approaches in standard radiographic image generation (x-ray) have been described (Kaplan, 1995). Scanning equalization radiography is a term used to describe a family of techniques designed to improve image quality, but systems that use this technique have limited commercial availability. Developments in the use of neural networks to decrease the number of false positive findings have been reported (Wu, 1995). Advances in radiology for use in lung cancer have focused largely on improving the image generating capabilities of CT (through the use of contrast media, reduced scanning time, and spiral volumetric technologies to improve image resolution) and developing other modalities such as MRI and ultrasonography to overcome the limitations of CT.

Recent advances in nuclear medicine instrumentation include multi-headed gamma cameras and higher efficiency computers for single-photon emission computed tomography (SPECT). New computer techniques in anatometabolic fusion imaging, which combine the images of CT, MRI, SPECT, or PET, are used to compare structural abnormalities with physiologic or metabolic information. These advances offer potential advantages over older procedures and other structural imaging techniques, not only in the diagnosis and staging of cancer, but also as tools for monitoring and predicting the effects of therapy on cancer biochemistry and metabolism.

Technologies designed to improve diagnostic accuracy are also being developed in areas other than imaging. In addition to the advances in early detection described previously, procedures and instrumentation used in sampling and analysis are being enhanced to increase diagnostic yield. Newer endoscopes with smaller diameters and greater flexibility have been developed in an attempt to improve access to and visualization of tumors. Local practices which use these other technologies may affect the utility of PET in the diagnostic work up.

Although advances in diagnostic imaging in lung cancer have focused mainly on improving earlier detection of disease and the accuracy of staging, the benefit of more accurate lung cancer staging has not been clearly demonstrated. The radiologic improvement of CT over standard x-ray is not disputed, and it has only been in recent years that CT has become sufficiently available and accepted as part of the routine diagnostic work up for lung cancer. However, the impact of CT on the reported prevalence of lung cancer and corresponding pre-test probability of disease has not been quantified. With respect to advances in staging using CT and alternatives such as PET, methods of reporting statistics on prevalence, natural history, and therapeutic effectiveness, and evaluations of diagnostic accuracy according to lesion size have been suggested to better define the impact of these advances (Black and Welch, 1993).

The effects of many therapeutic interventions for lung cancer are under investigation. The rationale for using PET in the clinical management of patients with an overall poor five year survival may be difficult to define. Any analysis of the effect of PET on the outcomes of treatment which might be attempted, based on other follow up of patients who have been reported in the existing literature, would be further complicated by the range of stages and histologies of non-small cell lung cancer included in the case series, and the associated range of treatment and outcomes.

V. SUGGESTIONS FOR FURTHER PET RESEARCH

Preliminary data have been published to date, and have attempted to define the operating characteristics of PET as a diagnostic test. Contributions from other investigators working with larger patient populations comparing PET to existing modalities will be needed to refine the characteristics of PET as a diagnostic tool in lung cancer, and to establish a base for further research.

In this context, future research within VA should focus on:

- establishing a PET registry, which would provide a range of data on demographic and clinical characteristics of patients on whom PET studies are performed, and on their clinical outcomes in a variety of settings;
- defining the role of PET as part of a diagnostic test battery;
- studies defining the impact of PET on treatment decision making and on outcomes such as survival, in comparison with existing technologies such as CT, MRI, and endoscopic procedures.

Table 3 Summary of the Literature: Diagnostic accuracy efficacy studies of PET and alternatives in lung cancer

Notes: All of the studies in the table are case series (Level V evidence) with internal controls (i.e. those with benign masses) used as a comparison group. All patients in these studies had suspected or biopsy-proven lung cancer (i.e. the pre-test probability of disease in the study populations was very high). Results from Knight, et al., 1996 and Inoue, et al., 1995 were reported as ranges to include data from all subgroup analyses.

None of these studies met strict evidence-based medicine criteria for blinding, but all studies presented information on blinding of the test interpreters to the biopsy gold standard. Blinding of the PET interpreters to other clinical and radiologic data varied across studies and is reflected in the columns designated "Operating Characteristics"; "PET + CT" indicates a complementary role of PET with CT, and PET alone indicates a substitutive role of PET for CT.

Where substantial duplication in purpose of study, patients studied, and results in multiple studies from the same institution could be inferred, only the most recent, largest, most rigorously designed, or most comprehensive was included in the table. Studies reviewed but not included are listed under "References".

Abbreviations are listed at the end of the table.

Role	***Study***	***N***	***Operating Characteristics****			***Evidence-Based Medicine Criteria*****			***Methodologic Quality Grade******
(Note: Some studies assessed multiple roles)			***PET***	***PET + CT***	***CT***	***comparison group***	***histologic gold standard***	***blinding***	
Defining unknown primary disease	Kubota, et al., 1990	12 malignant cases 10 benign cases		Se=83% Sp=90% accuracy=86%	no data reported	+ internal	+	+	**C**
	Scott, et al., 1994	47 malignant cases 15 benign cases		Se=94% Sp=80%	no data reported	+ internal	+	+	**C**
	Slosman, et al., 1994	31 malignant cases 5 benign cases		Se=93.5%	no data reported	+ internal	+ & follow-up	+	**C**
	Wahl, et al., 1994	19 malignant cases 4 benign cases	Se=100%		Se=100%	+ internal	+	+	**C**
	Sazon, et al., 1996	82 malignant cases 25 benign cases	Se=100% Sp=52%		no data reported	+ internal	+	+	**C**
	Knight, et al.,1996	32 malignant cases 16 benign cases		Se=100% Sp=58%-63% PPV=75% NPV=100%	Se=33%-41% Sp=52% PPV=83% NPV=52%	+ internal	+	+	**D**

Role (Note: Some studies assessed multiple roles)	Study	N	Operating Characteristics*			Evidence-Based Medicine Criteria**			Methodologic Quality Grade***
			PET	PET + CT	CT	comparison group	histologic gold standard	blinding	
Detecting overall lymph adenopathy	Patz, et al., 1995	42 patients with: 23 malignant nodes 39 benign nodes	Se=83% Sp=82%		Se=43% Sp=85%	+ internal	+	+	D
Detecting hilar/lobar lymph adenopathy	Patz, et al., 1995	42 patients with : 11 malignant nodes 29 benign nodes	Se=73% Sp=76%		Se=27% Sp=86%	+ internal	+	+	D
Detecting mediastinal lymph adenopathy	Patz, et al., 1995	42 patients with: 12 malignant nodes 10 benign nodes	Se=92% Sp=100%		Se=58% Sp=80%	+ internal	+	+	D
	Wahl, et al., 1994	23 patients with: 11 malignant sides 16 benign sides	Se=82% Sp=81% accuracy=81%		Se=64% Sp=44% accuracy=52%	+ internal	+	+	C
	Chin, et al., 1995	9 malignant cases 21 benign cases		Se=70% Sp=81% accuracy=80%	Se=56% Sp=86% accuracy=77%	+ internal	+	+	D
	Valk, et al., 1995	24 malignant sides 52 benign sides		Se=83% Sp=94% accuracy=91%	Se=63% Sp=73% accuracy=70%	+ internal	+ & follow-up	+	D
	Sazon, et al., 1996	32 patients with: 16 malignant sides 16 benign sides	Se=100% Sp=100%		Se=81% Sp=56%	+ internal	+	+	C
	Scott, et al., 1996	10 malignant nodes 65 negative nodes within: 9 malignant cases 18 benign cases		Se=100% Sp=98%-100%	Se=60% Sp=83%-94%	+ internal	+	+	D
Distinguishing local cancer recurrence from fibrosis	Patz, et al., 1994	35 recurrence cases 8 fibrosis cases		Se=97.1% Sp=100%	no data reported	+ internal	+ & follow-up	+	D
	Inoue, et al., 1995	23 recurrence cases 13 fibrosis cases		PET + x-ray, CT, MRI Se=100% Sp=56%-78% accuracy=86%	no x-ray, CT, or MRI data reported	+ internal	+ & follow-up	+	D

N, number of total study subjects included in analysis; unless otherwise noted, data are analyzed by subject
Se, sensitivity
Sp, specificity
PPV, positive predictive value
NPV, negative predictive value
CT, computed tomography
MRI, magnetic resonance imaging

** operating characteristics defined in Appendix 2: Assessing Diagnostic Technologies, Page 5-7*
*** Appendix 2, page 8*
**** Appendix 2, page 9*

Table 4 Diagnostic Accuracy Efficacy of PET in Lung Cancer

Notes: All of the studies in this table met most of the evidence-based criteria for diagnostic test evaluations. None of these studies met strict evidence-based medicine criteria for blinding, but all studies presented information on the comprehensiveness of blinding of test interpreters to the gold standard. All of the studies in the table are case series (Level V evidence); internal controls (i.e. those with benign masses) were used in each study, and it was possible to calculate sensitivity and specificity for PET in those studies. All patients in these studies had suspected or biopsy-proven lung cancer (i.e. the pre-test probability of disease in the study populations was very high); therefore, predictive values were not reported.

Where substantial duplication in purpose of study, patients studied, and results in multiple studies from the same institution could be inferred, only the most recent, largest, most rigorously designed, or most comprehensive was included in the table. Studies reviewed but not included are listed under "References".

Abbreviations are listed at the end of the table.

Study	***Patients/Methods***	***Results/Comments***
Kubota, et al., 1990 *(Tohoku University, Japan)*	**Purpose** to differentiate benign from malignant noncalcified lung tumors with PET using FDG and MET (MET data not reported) **Cases** 22 patients with unknown diagnosis presenting with a tumor shadow on chest x-ray **Methods** •all received CT before PET for anatomical placement •PET interpreted visually and with tumor/muscle radioactivity (TUR) ratios; cut-off 2.0 chosen prospectively to define lesions •image interpreters blinded to histology results •PET compared to histology **Limitations of study design** •small sample sizes •two PET scanners used in study •independent determination of test result and final diagnosis unclear	**Defining unknown primary disease (12 malignant cases, 10 benign cases)** •PET + CT: Se=83%; Sp=90%; accuracy=86% •TUR: malignant=4.4 ± 2.2 vs. benign=1.5 ± 0.3 ($p<0.001$) •CT: data not reported **Other findings** •using TUR tumors < 1 cm in diameter too small for accurate evaluation • one false positive attributed to granuloma •two false negatives attributed to a 0.5cm squamous cell carcinoma and liposarcoma •overall Se and Sp was higher in cases studied with Siemens model PT931/04 than with ECAT II model **Authors' Comments** while metabolic diagnosis allows for detection of cancer with PET, metabolic diagnosis is limited by heterogeneity of tumor metabolism

Study	Patients/Methods	Results/Comments
Patz, et al., 1994 *(Duke University, North Carolina)*	**Purpose** to assess PET in differentiating recurrent or residual bronchogenic carcinoma from fibrosis after therapy **Cases** 43 patients with a persistent radiographic abnormality after treatment for bronchogenic carcinoma •35 recurrences documented by pathology (n=25) or by clinical and radiographic progression (n=10) •8 with fibrosis **Methods** •all patients had CXR and CT interpreted prior to PET •chest x-ray and CT used to locate abnormality on PET •all PET scans conducted at least 2 months after completion of therapy, blinded to biopsy •ROI defined and SURs calculated and compared to biopsy •SUR threshold of 2.5 empirically determined to provide optimal Se and Sp for malignant disease **Limitations of study design** •prospective determination of SUR threshold unclear •number of cases and internal controls not equivalent (high prevalence of malignancy) •independent determination of test result and final diagnosis unclear	**Differentiating recurrence from fibrosis (35 recurrences, 8 fibrotic cases)** •PET + CT: Se=97.1% (95% CI=85.1%-99.9%); Sp=100% (95% CI=63.1%-100) •CT: data not reported **Semi-quantitative analysis [median (range)]** •SUR: recurrence= 7.7 (1.9-18.7) vs. fibrosis= 1.6 (0.6-2.4) (P=.0001) •34/35 patients with recurrence had SUR > 2.5 •8/8 with no evidence of recurrence 16-124 months after initial diagnosis; 6 biopsy-confirmed, 2 radiographically stable for at least 2 years after treatment **Authors' comments** •authors acknowledge small number of patients, but data suggest the usefulness of PET in differentiating recurrence from fibrosis in these patients •further validation of PET including a cost-benefit analysis requires a larger patient population
Inoue, et al., 1995 *(University of Texas, Houston, Texas)*	**Purpose** to evaluate the diagnostic accuracy of PET in detecting recurrent lung cancer **Cases** 38 patients with clinically suspected recurrent or residual lung cancer on conventional imaging (CI) (39 total lesions) •26 malignant, 13 benign **Methods** •all PET images interpreted visually in conjunction with CI (CT, MRI, or chest x-ray) blinded to biopsy •SUVs compared retrospectively in 25 patients •imaging compared to biopsy (n=11) or clinical/radiographic follow-up > 6 months after PET (n=28) **Limitations of study design** •SUV threshold determined retrospectively •independent determination of test result and final diagnosis unclear •number of cases and internal controls not equivalent •temporal differences between PET and clinical/radiologic follow-up	**Test characteristics based on visual inspection (23 recurrences, 13 fibrotic lesions)** •PET + CI: *Se=100%; *Sp=61.5%; *accuracy=86% •CI: data not reported •3 small cell cases not included **Semiquantitative analysis expressed as mean ± SD (16 recurrences, 9 benign cases)** •SUV: recurrence= 11.2 ± 5.7 vs. benign = 3.5 ± 1.8 (p < 0.0001) •no significant differences among patients with squamous cell, adenocarcinoma, or small cell histologies **Comparison of visual vs. quantitative PET results using SUV threshold of ≥ 5 defining non-small cell recurrence (13 recurrences, 9 benign cases)** •visual: *Se= 100%; *Sp= 55.5%; *accuracy= 82% •SUV: *Se= 100%; *Sp= 78%; *accuracy= 91% •3 small cell cases not included **Other findings** •PET false positives attributed to acute inflammation and reactive mesothelial cells •curvilinear pattern of FDG uptake noted in inflammatory lesions; focal uptake noted in recurrences; further study of FDG distribution is needed **Authors' comments** •PET should be interpreted in conjunction with anatomical imaging •further studies are needed to assess PET in distinguishing non-small cell and small cell cancers

Study	Patients/Methods	Results/Comments
Wahl, et al., 1994 *(Ann Arbor, Michigan and Orange Township Hospital, Australia)*	**Purpose** to prospectively evaluate the use of PET with FDG in mediastinal staging in patients with non-small cell lung cancer **Cases** 23 patients with abnormality on chest x-ray or strong suspicion of non-small cell lung cancer (i.e. high pre-test probability of cancer) who were to have mediastinal staging •27 mediastinal sides assessed (11 malignant, 16 benign) **Methods** •time interval between imaging studies 0-23 days, mean 1.9 days •diameter of primary lesions determined by CT; > 1 cm on short axis considered positive for mediastinal nodal involvement •SUV-lean calculated for primary lung lesion, expressed as mean ± SEM •chest x-ray available to locate primary lung lesion on PET •PET diagnosis of primary and mediastinal/hilar lymph reached by consensus using a four-point scale •level of certainty of concordance and discordance between CT and PET reached by consensus for fusion and nonfusion images •fusion imaging assessed for technical accuracy and rated on visual quality using a four-point scale •blinded, independent visual analysis of CT alone, PET alone, CT and PET together (nonfusion), and anatometabolic (fusion) images by two readers reached by consensus •biopsy confirmation determined by needle biopsy, direct observation, and surgery **Limitations of study design** •small sample size •number of cases and internal controls not equivalent •variations in nodal sampling techniques •independence of test result and determination of final diagnosis unclear • SUV cut-off for primary disease not reported *Note: Evaluation of diagnostic thinking efficacy (Level III) was attempted by authors when comparing diagnostic certainty of nonfused images and fusion images with CT and PET independently. However, results require histologic confirmation, and reconstruction of Se and Sp was not possible; therefore, inclusion of this study at a diagnostic efficacy level was felt to be warranted.*	**Defining unknown primary disease (19 malignant cases, 4 benign cases)** PET: Se=100% (n=23) CT: Se=100% (n=22) •PET SUV-lean: malignant= 6.82 ± 0.983 vs. benign= 1.047 ± 0.268 (*P* < .04) •size on CT: malignant= 34.9mm ± 2.6 vs. benign=15.5mm ± 2.2 (*P* < .005) **Mediastinal/Hilar nodal disease (11 malignant sides, 16 benign sides)** •PET: Se=82%; Sp=81%; accuracy=81% •CT: Se=64%; Sp=44%; accuracy=52% •2 patients with hilar involvement not included in calculations •CT and PET false negatives attributed to close proximity of hilar and mediastinal nodes •PET and CT false positives attributed to granulomas, anthracotic disease; one on PET due to hilar proximity; seven on CT attributed to enlarged nodes **Combined nonfused CT and PET images** •CT + PET judged better than PET alone in 7/22 cases •CT + PET judged better than CT alone in 16/22 cases •one case not included in PET scan field of view **PET anatometabolic fusion images (histologic proof for 7 cases)** •visual or fusion images changed overall CT interpretation in 16/22 patients •on CT 3 negative node cases were changed to positive; in 11 node cases were changed from positive to negative •PET results correlated with all histologically confirmed cases •on CT 3 cases changed from tumor to nontumor on fusion image •one positive PET represented atelectasis on fusion image •one false positive fusion image due to mediastinal invasion not confirmed at surgery **Other findings** •experience too limited to comment on diagnosing hilar nodal involvement •additional data from larger studies needed to confirm results and demonstrate effect on treatment planning and patient outcomes •PET alone or with CT is preferred approach for noninvasive staging of metastatic mediastinal lymph nodes in patients with newly diagnosed or suspected non-small cell lung cancer

Study	Patients/Methods	Results/Comments
Scott, *et al.*, 1994 *(Creighton University Medical Center and Omaha Veterans Affairs Medical Center, Omaha, Nebraska)*	**Purpose** •to retrospectively compare the accuracy of PET and CT to CT alone in imaging hilar and mediastinal lymph nodes •to define the initial experience with PET imaging in patients with various lung tumors **Cases** 62 patients with various lung abnormalities •all patients had biopsy confirmed primary disease (47 malignant, 15 benign) •25 patients had biopsy confirmation of mediastinal lymph nodes involvement (3 malignant, 22 benign) **Methods** •all patients underwent CT before PET; CT or chest x-ray data used to locate lung mass on PET for visual analysis, but not always available for DUR calculations •abnormal mediastinal node defined as > 1 cm in diameter on CT •PET images analyzed visually by one reader, DURs calculated •mean DURs for benign and malignant primary tumors compared, expressed as mean ± SEM •PET images blinded to biopsy •PET + CT and CT alone compared to biopsy results **Limitations of study design** •limited histologic data available on mediastinal lymph nodes •retrospective design •number of cases and internal controls not equivalent	**Detecting unknown primary disease (47 malignant cases, 15 benign cases)** •PET + CT: Se=93.6%; Sp=80% •CT: no data reported **Mediastinal lymph nodes (3 malignant cases, 22 benign cases)** •PET correctly identified 19/22 benign cases •PET correctly identified 2/3 malignant cases •PET identified subcarinal lymph node metastases in 1 of 2 malignant cases with normal CT •CT correctly identified 20/22 benign cases •CT correctly identified 1/3 malignant cases **Quantitative analysis of PET for primary tumors (expressed as mean DUR ± SEM)** •benign= 1.14 ± 0.26 vs. malignant=6.4 ± 0.56 ($p < 0.0001$) •in retrospect, DUR cutpoint of 2.0 produced greatest accuracy (92%): *Se= 94%; *Sp=87% **Other findings** •PET false positives attributed to granuloma or inflammatory disease •PET false negatives attributed to small tumors < 1cm^2 and low grade malignancy **Authors' comments** •no false negative PET findings occurred in patients with elevated glucose levels •confirmation from larger series is needed •PET most useful as adjunct to CT •evaluation of PET in measuring treatment response is needed
Patz, et al., 1995 *(Duke University, North Carolina)*	**Purpose** to assess prospectively the diagnostic accuracy of PET in detecting thoracic lymph node metastases in patients with bronchogenic carcinoma **Cases** 42 patients with untreated bronchogenic carcinoma determined by chest x-ray, CT, bone scan and PET, who were to have nodal sampling •40 non-small cell; 1 small cell type; 1 undifferentiated type 62 total nodal stations sampled at surgery •40 hilar/lobar nodes; 22 mediastinal nodes **Methods** •thoracic CT performed before PET •lymph nodes on CT > 1 cm on short axis diameter classified as abnormal •CT and qualitative PET results read independently and blinded to biopsy •CT, PET, and surgical stage mapped according to ATS classification system •limited, partial histologic sampling done in 40 patients; 2 patients had single nodal station sampled by thin-needle aspiration •PET and CT scans compared to biopsy **Limitations of study design** •source of cohort influenced by test results •variations in comprehensiveness of nodal sampling •test result and determination of final diagnosis not independent	**Detecting hilar/lobar lymph node metastases (11 metastatic nodes, 29 benign nodes)** PET: Se=73%; Sp=76% (p=0.009 PET compared with pathology) CT: Se=27%; Sp=86% (p=0.369 CT compared with pathology) **Detecting metastatic lymph node metastases (12 metastatic nodes, 10 benign nodes)** PET: Se=92%; Sp=100% (p<0.001 PET compared with pathology) CT: Se=58%; Sp=80% (p=0.099) CT compared with pathology **Overall detection (23 metastatic nodes, 39 benign nodes)** PET: Se=83%; Sp=82% (p<0.001 PET compared with pathology) CT: Se=43%; Sp=85% (p=0.019 CT compared with pathology) **Other findings** •false positive hilar nodes on PET due to inflammatory response

Study	*Patients/Methods*	*Results/Comments*
Chin, et al., 1995 *(Bowman Gray School of Medicine, Wake Forest University, North Carolina)*	**Purpose** to assess prospectively the role of PET in evaluating mediastinal nodal metastases in patients with non-small cell lung cancer **Cases** •30 patients with potentially resectable tumors (N0-N1 and N2 disease) determined by CT **Methods** •lymph node on CT considered positive if long-axis diameter > 1.5 cm •CT and qualitative PET assessed independently according to ATS classification; results based on presence or absence of disease in mediastinum •SUVs calculated, but values not used to compute results •surgeons aware of clinical, radiologic, and PET results for mediastinal exploration; ipsilateral mediastinal explorations performed in all patients; contralateral explorations not routinely performed in absence of radiologic evidence •imaging results compared to biopsy **Limitations of study design** •number of cases and internal controls not equivalent •comprehensiveness of nodal sampling not noted •test result and determination of final diagnosis not independent for mediastinal evaluations	**Detecting primary tumor (39 malignant cases, 0 benign cases)** PET + CT: Se=94%; accuracy=89% CT: data not reported •data with which to perform calculations were not presented **Detecting metastatic lymph node metastases (9 cases, 21 controls)** PET + CT: Se=70%; Sp=81%; accuracy=80% CT: Se=56%; Sp=86%; accuracy=77% •agreement between CT and PET in 21 patients (70%) with a diagnostic accuracy of 90%; correlation between combined images and surgical results was statistically significant (p=0.004) **Authors' comments** •low Se of CT was a function of rigorous preoperative evaluation, limiting the number of true positives detected by either imaging modality and introducing a bias against CT •because of the number of false positive and false negative results, PET should not supplant histologic confirmation •lack of precise correlation of nodal stations between surgical results and both imaging modalities is unlikely to affect clinical management or outcome •low resolution of PET affects its ability to distinguish mediastinal tumors, but may be overcome by coregistration with CT •PET may contribute best in those patients whose CT image shows normal mediastinal adenopathy despite a high index of suspicion of N2 disease, persons with high operative risks, or low-risk patients whose lymph nodes meet size criteria (> 1 cm but < 2 cm) on CT, and may direct attention toward previously unsuspected areas of disease •role of PET may be influenced by local practices particularly with respect to the routine use of mediastinoscopy prior to surgery
Valk, et al., 1995 (*Northern California PET Imaging Center and Radiologic Associates of Sacramento, Sacramento, California)*	**Purpose** to assess prospective the role of PET with CT in staging patients with suspected or known lung cancer **Cases** 74 patients referred to the PET Center for staging of histologically diagnosed non-small cell lung cancer: •76 total mediastinal sides with histologic confirmation (24 positive, 52 negative) •7 patients with hilar node involvement •18 patients with distant metastases **Methods** •CT performed at referring site •criteria for positive nodes on CT defined as > 1 cm on short axis diameter •PET images interpreted visually and quantitatively; CT images used for localization and measurement of primary tumor, mediastinal CT findings were disregarded •SUVs calculated for primary lesions > 2 cm diameter •imaging correlated to biopsy obtained at mediastinoscopy or thoracotomy or to follow-up; analyzed by mediastinal side •effect of experience in PET interpretation during study evaluated; graded on a three-point visual scale •distant metastases determined by biopsy (n=6) or clinical follow up (n=19) **Limitations of study design** •number of cases and internal controls not equivalent (high index of suspicion for lung cancer) •test result and determination of final diagnosis not independent •comprehensiveness of nodal sampling unclear	**Detecting primary tumor** •authors reported a relative lack of anatomic information with PET •no operating characteristics reported for either PET or CT **Detecting mediastinal lymph node involvement (24 positive sides, 52 negative sides)** PET + CT: Se=83%; Sp=94%; accuracy=91% CT: Se=63%; Sp=73%; accuracy=70% **Detecting distant metastases** •PET showed evidence of distant metastases in 18 (18%) patients; only 12 patients had histologic or clinical confirmation, with no false positive PET results •authors reported more false positive distant CT findings in 19 patients (19%) than true positive **Other findings** •mean interval between CT and PET scans=23 days (range 1-51 days) •PET correctly changed the N stage as determined by CT in 18 (24%) staging evaluations; in 16 of 17 discordant cases, PET proved to be correct •false negative PET results attributed to micrometastasis; PET limited by minimal detectable tumor mass •false positive PET results attributed to anthracotic lymph nodes and hyperplasia

Study	Patients/Methods	Results/Comments
Sazon, et al., 1996 *(Veterans Affairs Medical Center, West Los Angeles, California)*	**Purpose** to assess the diagnostic accuracy of PET and CT in detecting and staging lung cancer **Cases** 107 patients with an abnormal chest x-ray: 73 with non-small cell; 5 small cell; 25 various benign chest diseases •of 73 with non-small cell: 32 had mediastinal evaluation (16 malignant, 16 benign) •4 distant metastases **Methods** •PET and CT read independently •PET results for primary lung lesion and mediastinal involvement determined qualitatively •criteria for positive nodes on CT defined as nodal enlargement > 1 cm in diameter on transaxial images •mediastinal evaluation accomplished by transbronchial needle biopsy, mediastinoscopy, thoracotomy, or at autopsy •PET and CT interpretations blinded to and correlated with biopsy **Limitations of study design** •number of cases and internal controls not equivalent •influence of PET result on choice of patient cohort for mediastinal evaluation unclear •variations in nodal sampling techniques •independence of test result and determination of final diagnosis unclear	**Detecting primary disease (82 malignant cases, 25 benign cases)** PET: Se=100%; Sp=52% CT: data not reported **Detecting mediastinal lymph node involvement (16 malignant cases, 16 benign cases)** PET: Se=100%; CT: Se=81% (95% CI of difference=-13% to 39.3%; p=0.22) PET: Sp=100%; CT: Sp=56% (95% CI of difference= 15.3% to 72.7%; p=0.01) **Authors' comments** •low Sp of PET in detecting primary disease may be due to broader range of patients chosen for study and to criteria used for PET scan results •authors report the potential of PET in mediastinal staging and distant metastases with whole body imaging requires further confirmation
Scott, et al., 1996 *(Creighton University and Omaha VAMC, Omaha, Nebraska)*	**Purpose** to prospectively assess the role of PET and CT versus CT alone in detecting N2 or N3 lymph node metastases **Cases** •27 patients with CT evidence of known or suspected NSCLC; 75 total lymph node stations analyzed •exclusions included: - patients not appropriate for mediastinoscopy or thoracotomy, or - patients with solitary pulmonary nodules ≤ 2 cm in diameter without evidence on CT of mediastinal lymph node involvement **Methods** •all patients underwent CT, PET and surgical staging •CT results used to determine location of lung mass or regional metastases on PET •CT and PET scans read by separate radiologists blinded to surgical staging results •ATS lymph node mapping system used to correlate nodes on imaging with biopsy; biopsy procedures included scalene node biopsy, mediastinoscopy, or thoracotomy •both CT and PET images available to surgeon during the operation •CT criteria for positive lymph node was 1.0 cm in short-axis diameter •PET criteria for positive lymph nodes was a SUV > 4.2; for lung masses was > 2.0 **Limitations of study design** •number of cases and internal controls not equivalent •test result and determination of final diagnosis not independent •variations in nodal sampling technique •influence of PET results on choice of patient cohort unclear	**Detecting mediastinal lymph node involvement by patient (9 malignant cases, 18 benign cases)** PET + CT: Se=100%; Sp=100% CT: Se=67%; Sp=83% •CT had 3 false positive and 3 false negative diagnoses **Detecting mediastinal lymph node involvement by node (10 positive nodes, 65 negative nodes)** PET + CT: Se=100%; Sp=98% CT: Se=60%: Sp=94% •PET had 1 false positive; CT had 4 false positives and 4 false negatives **Other findings** •when data were analyzed by patient, there were 6 discrepencies between CT and PET over the presence or absence of positive lymph nodes in the mediastinum (p=0.031) •when data were analyzed by node, there were 9 discrepencies between CT and PET over the presence or absence of positive lymph nodes in the mediastinum (p=0.039) **Authors' comments** •more data are required to determine the optimal threshold value for mediastinal evaluations

Study	Patients/Methods	Results/Comments
Knight, et al., 1996 *(Vanderbilt University, Nashville, Tenneessee)*	**Purpose** to prospectively assess PET in differentiating benign from malignant primary lung cancer in patients with and without a history of malignancy **Cases** 48 patients with lesions suspicious for malignancy on chest x-ray and CT: •Group I- 27 patients with no prior history of malignancy (15 malignant, 12 benign) •Group 2- 19 patients with history of malignancy (17 malignant, 4 benign) **Methods** •initial chest x-ray and CT interpretations blinded to PET results •all patients fasted before PET •independent PET interpretations blinded to CT and chest x-ray interpretations, but chest x-ray and CT data used to locate lesion on PET •ROIs determined; SUR and L/B ratios calculated; SUR cut-off > 2.5 and L/B cut-off > 5 chosen to define malignant nodules •SURs not obtained due to motion artifact (n=1) and partial volume effect (n=2)for lesions < 1 cm •confirmation by various biopsy procedures (n=30); by pleural cytology (n=2); clinical and radiologic follow-up (n=16) for 6-16 months •radiologic confirmation of malignant disease defined as increase in size on follow-up CT performed between 6-12 months after initial evaluation; size parameter not defined •PET + CT vs. CT compared with biopsy or follow-up, reported by patient **Limitations of study design** •except in Group 1, number of cases and internal controls not equivalent (high prevalence of malignancy) •PET result and determination of final diagnosis not independent •temporal differences between PET scans and clinical/radiologic follow-up	**Semiquantitative analysis of unknown primary (reported as mean ± SD)** Group 1 (14 malignant cases, 12 benign cases) SUR: malignant= 8.9 ± 4.9 vs. benign= 3.3 ± 3.2 (p = 0.001) L/B: malignant= 20.6 ± 14.2 vs. benign= 5.2 ± 5.5 (p = 0.0008) Group 2 (15 malignant cases, 4 benign cases) SUR: malignant= 8.9 ± 5.1 vs. benign= 1.3 ± 1.0 (p = 0.00003) L/B: malignant=13.0 ± 8.3 vs. benign= 2.6 ± 3.2 (p = 0.0009) Group 1 + 2 (29 malignant cases, 16 benign cases) SUR: malignant=8.9 ± 5.0 vs. benign= 2.8 ± 2.9 (p = 0.00003) L/B: malignant=16.6 ± 12.0 vs. benign= 4.5 ± 5.0 (p = 0.00001) **Corresponding operating characteristics (for lesions > 1 cm)** Group 1 (15 malignant, 12 benign) PET + CT: Se=100%; Sp=58%; PPV=75%; NPV=100% CT: Se=33%*; Sp=52%*; PPV=83%*; NPV=52%* Group 2 (17 malignant, 4 benign) PET + CT: Se=100% CT: Se=41%* Group 1 + 2 (32 malignant, 16 benign) PET + CT: Se=100%; Sp=62.5% CT: Se=38%*; Sp=88%* **Other findings** •findings using SUR correlated with findings using L/B ratio in distinguishing benign from malignant disease •six false-positives due to tuberculosis, granuloma, schwannoma, fibrous mesothelioma with focally increased cellularity, and inflammatory mass with macrophages **Authors' comments** •study is limited by patient selection bias

Abbreviations: Se,sensitivity
Sp,specificity
CT,computerized tomography
ROI,region of interest
SUV,standard uptake value
L/B, lesion-to-background ratio
DUR,differential uptake ratio
ATS, American Thoracic Society

*indicated calculated by MDRC TA Program from data supplied in published article

Table 5 Hypothetical Diagnostic Thinking Efficacy of PET in Lung Cancer

Notes *The study in this table met most of the evidence-based criteria for diagnostic test evaluations and is a case series (Level V evidence); all of the patients presented with a high index of suspicion for lung cancer. Although internal controls (i.e. those with benign masses) were used, an insufficient number was available to calculate specificity. Calculations using Bayesian analysis are hypothetical and are used for illustrative purposes.*

Abbreviations are listed at the end of the table.

Study	*Patients/Methods*	*Results/Comments*
Slosman, et al., 1993 (Geneva University Hospital, Switzerland)	**Purpose** • to measure the sensitivity of PET in detecting lung cancer • to determine prospectively the role of PET scanning in a satellite center as an adjunct to conventional methods using Bayesian analysis **Cases** 36 patients presented to center with suspected lung cancer of various types and stages based on x-ray CT and clinical work-up, and who were scheduled for thoracotomy •21 patients with histological proof of cancer at time of the PET scan •15 patients with a pulmonary mass of unknown etiology **Methods** •all patients received CT and PET before treatment •ROIs calculated; FDG uptake expressed as tumor to non-tumor ratio (TNT) •positive PET scan defined prospectively as TNT ratio ≥ 1.5 •33 patients obtained final diagnosis by thoracotomy; 3 by observation •Bayesian analysis performed using Se from this study and Sp from Kubota, et al., 1990 **Limitations of study design** •number of cases and internal controls not equivalent (high prevalence of malignancy) •independence of test result and determination of final diagnosis unclear	**Defining unknown primary disease (31 malignant cases, 5 benign cases)** •PET: Se=93.5% •unable to calculate specificity due to low number of be not reported •CT: data not reported **Bayesian Analysis (based on Se=93.5% and Sp=75%)** (pD+=prevalence of disease in a population) •pD+=.80, PET scan (-), the post test probability of disease=26% •pD+=.80, PET scan (+), the post test probability=97% •pD+=.20, PET scan (-), the post test probability of disease=2% •pD+=.20, PET scan (+), the post test probability of disease=29% **Other findings** •two false positives due to significant inflammation •two false negatives due to small tumor size and/or limitations in spatial resolution **Authors' comments** •PET has greatest impact in a population with a low prevalence of cancer and in whom a positive test could lead to more aggressive therapy than one would expect otherwise •PET appears of little use in avoiding a thoracotomy in a patient with a high prevalence of cancer •choice of the appropriate TNT ratio threshold needs further study

Abbreviations: Se=sensitivity
Sp=specificity
CT=computerized tomography
ROI=region of interest

*indicated calculated by MDRC TA Program from data supplied in published article

VI. REFERENCES Background and studies meeting evidence-based medicine criteria for evaluations of diagnostic tests

American Cancer Society. *Cancer Facts & Figures-1996*. New York: National Media Office-ACS, 1996.

Beckett WS. Epidemiology and Etiology of Lung Cancer. *Clinics in Chest Medicine* 1993; 14(1):1-15.

Black W and Welch HG. Advances in Diagnostic Imaging and Overestimations of Disease Prevalence and the Benefits of Therapy. *New England Journal of Medicine* 1993; 328(17):1237-43.

Buccheri G and Ferrigno D. Prognostic factors in lung cancer: tables and comments. *European Respiratory Journal* 1994;7:1350-64.

Bunn PA. Future directions in clinical research for lung cancer. *Chest* 1994;106(6):399s-407s.

Chin R, Ward R, Keyes JW, Choplin RH, Reed JC, Wallenhaupt, *et al.* Mediastinal staging of non-small-cell lung cancer with positron emission tomography. *American Journal of Respiratory and Critical Care Medicine* 1995;152:2090-6.

Chiti A, Maffioli LS, Infante M, Grasselli G, Incarbone M, Gasparini MD, et al. Assessment of mediastinal involvement in lung cancer with technetium-99m-sestamibi SPECT. *The Journal of Nuclear Medicine* 1996;37:938-42.

Colice GL, Birkmeyer JD, Black WC, Littenberg B, Silvestri G. Cost-effectiveness of head CT in patients with lung cancer without clinical evidence of metastases. *Chest* 1995;108:1264-71.

Dales RE, Stark RM, Raman S. Computed tomography to stage lung cancer. *American Review of Respiratory Disease* 1990;141:1096-1101.

Filderman AE and Matthay RA. Bronchogenic carcinoma. In: Richard A. Matthay, ed. *Pulmonary and Critical Medicine Volume 1.* St. Louis: Mosby-Year Book, Inc., 1994:1-17.

Gloeckler Ries LA. Influence of extent of disease, histology, and demographic factors on lung cancer survival in the SEER population-based data. *Seminars in Surgical Oncology* 1994; 10:21-30.

Greenes RA and Brinkley JF. Radiology systems. In: *Medical Informatics- Computer Applications in Health Car*e. Reading, Massachusetts; Addison-Wesley Publishing Co., 1990.

Gupta NC and Frick M. Clinical applications of positron emission tomography in cancer. *CA-A Cancer Journal for Clinicians* 1993;43:235-54.

Guyatt GH, Lefcoe M, Walter S, Cook D, Troyan S, Griffith L, *et. al.* Interobserver variation in the computed tomographic evaluation of mediastinal lymph node size in patients with potentially resectable lung cancer. *Chest* 1995;107:116-9.

Hennekens CH, Buring JE, Manson, JE, Stampfer M, Rosner, B, Cook NR, *et al.* Lack of effect of long-term supplementation with beta carotene on the incidence of malignant neoplasms and cardiovascular disease. *New England Journal of Medicine* 1996;334:1145-9.

Hillers TK, Sauve MD, Guyatt GH. Analysis of published studies on the detection of extrathoracic metastases in patients presumed to have operable non-small cell lung cancer. *Thorax* 1994;49:14-19.

Gambhir SS, Hoh CK, Phelps ME, Madar I, Maddahi J. Decision tree sensitivity analysis for cost-effectiveness of FDG-PET in the staging and management of non-small-cell lung carcinoma. *The Journal of Nuclear Medicine* 1996;37(9):1428-1436.

Inoue T, Kim EE, Komaki R, Wong FCL, Bassa P, Wong W, *et al*. Detecting recurrent or residual lung cancer with FDG-PET. *Journal of Nuclear Medicine* 1995;36:788-93.

Kabat GC. Recent developments in the epidemiology of lung cancer. *Seminars in Surgical Oncology* 1993;9:73-9.

Kaplan D and Goldstraw P. New techniques in the diagnosis and staging of lung cancer. *Cancer Treatment and Research* 1995;72;223-54.

Kirby TJ, Mack MJ, Landreneau RJ, Rice TW. Lobectomy--video-assisted thoracic surgery versus muscle-sparing thoracotomy: a randomized trial. *The Journal of Thoracic and Cardiovascular Surgery* May 1995;109:997-1002.

Knight SB, Delbeke D, Stewart JR, Sandler MP. Evaluation of pulmonary lesions with FDG-PET: Comparison of findings in patients with and without a history of prior malignancy. *Chest* 1996;109:982-8.

Kubota K, Matsuzawa T, Fujiwara T, Ito M, Hatazawa J, Ishiwata K, *et al*. Differential diagnosis of lung tumor with positron emission tomography: a prospective study. *Journal of Nuclear Medicine* 1990;31:1927-33.

Lederle FA, Niewoehner DE. Lung cancer surgery: a critical review of the evidence. *Archives of Internal Medicine* 1994;154:2397-2400.

Minna J . Neoplasms of the lung. In: Isselbacher K, *et al*., eds. *Harrison's Principles of Internal Medicine*. 13th ed. New York: McGraw-Hill, Inc., 1994:1221-9.

Mountain CF. Lung cancer staging classification. *Clinics in Chest Medicine* 1993;14(1):43-51.

Nieweg OE. Potential applications of positron emission tomography in surgical oncology. *European Journal of Surgical Oncology* 1994;20:415-24.

Nolop K, Rhodes CG, Brudin LH, Beaney RP, Krausz T, Jones T, *et al.* Glucose utilization in vivo by human pulmonary neoplasms. *Cancer* 1987;60:2682-9.

Office of the Secretary of Veterans Affairs. *Annual Report of the Secretary of Veterans Affairs: Fiscal Year 1994* Washington DC: U.S. Government Printing Office, 1995:54-5.

Panzer R, Black E, Griner P. *Diagnostic Strategies for Common Medical Problems* Philadelphia, PA: American College of Physicians, 1991:16-28.

Patz EF, Lowe VJ, Hoffman JM, Paine SS, Harris LK, Goodman PC. Persistent or recurrent bronchogenic carcinoma: detection with PET and 2-[F-18]-2-deoxy-D-glucose. *Radiology* 1994; 191:379-82.

Patz EF, Lowe VJ, Goodman PC, Herndon J. Thoracic nodal staging with pet imaging with ^{18}FDG in patients with bronchogenic carcinoma. *Chest* 1995;108:1617-21.

Pearson F. Staging of the mediastinum: role of mediastinoscopy and computed tomography. *Chest* 1993; 103(4):346s-8s.

Pugatch RD. Radiologic evaluation in chest malignancies. *Chest* 1995;107:2945-75.

Quint LE, Francis IR, Wahl RL, Gross BH, Glazer GM. Preoperative staging of non-small-cell carcinoma of the lung: imaging methods. *American Journal of Roentgenology* 1995;164:1349-59.

Samet JM. The epidemiology of lung cancer. *Chest* 1993;103(1):20S-29S.

Sazon DAD, Santiago SM, Soo Hoo GW, Khonsary A, Brown C, Mandelkern M, *et al.* Fluorodeoxyglucose-positron emission tomography in the detection and staging of lung cancer. *American Journal of Respiratory and Critical Care Medicine* 1996;153:417-21.

Scott W , Schwabe JL, Gupta NC, Dewan NA, Reeb SD, Sugimoto JT, *et al.* Positron emission tomography of lung tumors and mediastinal lymph nodes using F-18-fluorodeoxyglucose. *Annals of Thoracic Surgery* 1994;58:698-703.

Scott WJ, Gobar LS, Terry JD, Dewan NA, Sunderland JJ. Mediastinal lymph node staging of non-small-cell lung cancer: a prospective comparison of computed tomography and positron emission tomography. *The Journal of Thoracic and Cardiovascular Surgery* 1996;111:642-8.

Seely JM, Mayo JR, Miller RR, Müller NL. T1 lung cancer: prevalence of mediastinal nodal metastases and diagnostic accuracy of CT. *Radiology* 1993;186:129-32.

Shepherd F . Screening, diagnosis, and staging of lung cancer. *Current Opinion in Oncology* 1993; 5:310-22.

Shepherd F. Treatment of advanced non-small cell lung cancer. *Seminars in Oncolology* 1994; 21(4 Supplement 7):7-18.

Schomburg A, Bender H, Reichel C, Sommer T, Ruhlman J, Kozak B, et al. Standardized uptake values of fluorine-18 fluorodeoxyglucose: the value of different normalization procedures. *European Journal of Nuclear Medicine* 1996;23(5):571-4.

Silvestri GA, Littenberg B, Colice GL. The clinical evaluation for detecting metastatic lung cancer: a meta-analysis. *American Journal of Respiratory and Critical Care Medicine* 1995;152:225-30.

Slosman DO, Spiliopoulos A, Couson F, Nicod L, Louis O, Lemoine R, *et al.* Satellite PET and lung cancer: a prospective study in surgical patients. *Nuclear Medicine Communications* 1993; 14:955-61.

Souquet PJ, Chauvin F, Boissel JP, Cellerino R, Cormier Y, Ganz PA, *et al.* Polychemotherapy in advanced non small cell lung cancer: a meta-analysis. *Lancet* 1993;342:19-21.

Strauss GM, Kwiatkowski DJ, Harpole DH, Lynch TJ, Skarin AT, Sugarbaker DJ. Molecular and pathologic markers in stage I non-small-cell carcinoma of the lung. *Journal of Clinical Oncology* 1995;13:1265-79.

Sugarbaker DJ, Strauss GM. Advances in surgical staging and therapy of non-small-cell lung cancer. *Seminars in Oncology* 1993;20(2):163-72.

Szabo E, Birrer MJ, Mulshine JL. Early detection of lung cancer. *Seminars in Oncology* 1993; 20(4):374-82.

Tockman MS, Bupta PK, Pressman NJ, Mulshine JL. Considerations in bringing a cancer biomarker to clinical application. *Cancer Research* 1992; 52(Supplement):2711s-8s.

Valk PE, Pounds TR, Hopkins DM, Haseman MK, Hofer GA, Greiss HB, *et al.* Staging non-small cell lung cancer by whole-body positron emission tomographic imaging. *Annals of Thoracic Surgery* 1995;60:1573-82.

Wahl R , Quint LE, Greenough RL, Meyer CR, White RI, Orringer MB. Staging of mediastinal non-small cell lung cancer with FDG PET, CT, and fusion images: preliminary prospective evaluation. *Radiology* 1994;191:371-7.

Webb WR, Sarin M, Zerhouni EA, Heelan RT, Glazer GM, Gatsonis C. Interobserver variability in CT and MR staging of lung cancer. *Journal of Computer Assisted Tomography* 1993; 17(6):841-6.

Wiedemann HP, Meziane M . Diagnostic imaging. In: Richard A. Matthay, ed. *Pulmonary and Critical Care Medicine*. St. Louis: Mosby-Year Book, Inc., 1994:1-26.

Wu YC, Doi K, Giger ML. Detection of lung nodules in digital chest radiographs using artificial neural networks: A Pilot Study. *Journal of Digital Imaging* 1995;8(2):88-94.

Zagonel V, Pinto A, Serraino D, Babare R, Serraino D, Babare R, *et al.* Lung cancer in the elderly. *Cancer Treatment Reviews* 1994;20:315-29.

VII.REFERENCES: Technical efficacy studies (not included in tables)

Bury T, Dowlati A, Paulus P, Hustinx R, Radermecker M, Rigo P. Staging of non-small-cell lung cancer by whole-body fluorine-18 deoxyglucose positron emission tomography. *European Journal of Nuclear Medicine* 1996;23(2);204-6.

Frank A, Lefkowitz D, Jaeger S, Gobar L, Sunderland J, Gupta N, Scott W, *et al.*. Decision logic for retreatment of asymptomatic lung cancer recurrence based on positron emission tomography findings. *International Journal of Oncology, Biology, Physics* 1995;32(5):1495-1512.

Hebert ME, Lowe VJ, Hoffman JM, Patz EF, Anscher MS. Positron emission tomography in the pretreatment evaluation and follow-up of non-small cell lung cancer patients treated with radiotherapy: preliminary findings. *American Journal of Clinical Oncology*. 1996;19(4):416-21.

Hunter GJ, Hamberg LM, Alpert NM, Choi NC, Fischman AJ. Simplified measurement of deoxyglucose utilization rate. *The Journal of Nuclear Medicine*. 1996;37(6):950-5.

Hübner KF, Buonocore E, Singh SK, Gould HR, Cotten DW. Characterization of chest masses by FDG positron emission tomography. *Clinical Nuclear Medicine* 1995;20:293-8.

Lewis P, Griffin S, Marsden P, Gee T, Nunan T, Malsey M, Dussek J. Whole-body ^{18}F-fluorodeoxyglucose positron emission tomography in preoperative evaluation of lung cancer. *Lancet* 1994;344:1265-6.

Miyauchi T, Wahl RL. Regional 2-[^{18}F]fluoro-2-deoxy-D-glucose uptake varies in normal lung. *European Journal of Nuclear Medicine*. 1996;23:517-23.

Rege SD, Hoh CK, Glaspy JA, Aberle DR, Dahlbom M, Razavi MK. Imaging of pulmonary mass lesions with whole body positron emission tomography and fluorodeoxyglucose. *Cancer* 1993; 72:82-90.

Sasaki M, Ichiya Y, Kuwabara Y, Akashi Y, Yoshida T, Fukumura T, et al. The usefulness of FDG positron emission tomography for the detection of mediastinal lymph node metastases in patients with non-small cell lung cancer: a comparative study with x-ray computed tomography. *European Journal of Nuclear Medicine*. 1996;23(7):741-7.

VIII. REFERENCES: Excluded studies

Exclusion criteria included:

- number of cases < 12
- duplicated or superseded by subsequent or concurrent study from the same institution
- abstract, not peer reviewed

Abe Y, Matsuzawa T, Fujiwara T, Itoh M, Fukuda H, Yamaguchi K, *et al.* Clinical assessment of therapeutic effects on cancer using 18-F-2-Fluoro-2-deoxy-d-glucose and positron emission tomography: preliminary study of lung cancer. *International Journal of Radiology, Oncology, Biology, Physics* 1990;19:1005-10.

Gupta NC, Boman BM, Frank AR, Sunderland J, Shiue CY, Frick MP. Utility of PET-FDG imaging in treatment planning and monitoring of lung tumors. *Radiology* 1991;181:152. (abstract)

Hughes JMB. ^{18}F-fluorodeoxyglucose PET scans in lung cancer. *Thorax.* 1996;51(Suppl 2):S16-22. (Review)

Hoh C, Hawkins RA, Glaspy JA, Dahlbom M, Tse NY, Hoffman EJ, *et al.* Cancer detection with whole-body PET using 2-F18-fluoro-2-deoxy-D-glucose. *Journal of Computer Assisted Tomography* 1993;17(4):582-9.

Ichiya Y, Kuwabara Y, Otsuka M, Tahara T, Yoshikai T, Fukumura T, *et al* Assessment of response to cancer therapy using fluorine-18-fluorodeoxyglucose and positron emission tomography. *Journal of Nuclear Medicine* 1991;32:1655-60.

Kim E, Chung S, Haynie TP, Chang-Guen K, Byung-Jae C, Podoloff DA, *et al.* Differentiation of residual or recurrent tumors from post-treatment changes with F-18 FDG PET. *RadioGraphics* 1992;12:269-79.

Knopp MV, Strauss LG, Haberkorn U, Dimitrakopoulu A, Bischoff H, Branscheid D, *et al.* Positron emission tomography with F-18-deoxyglucose in the imaging and staging of bronchogenic carcinoma. *European Journal of Nuclear Medicine* 1990;16:560 (abstract).

Knopp MV, Strauss LG, Bischoff H, Doll JK, Vogt-Moykopf I, Maier-Borst W, *et al.* PET imaging in the evaluation of recurrent thoracic tumors. *Radiology* 1991;181:152 (abstract).

Knopp MV, Strauss LG, Dimitrakopoulu A, Haberkorn UA, Blatter J, van Kaick G. Use of PET for optimization of chemotherapy of lung malignancies. *Radiology* 1989;173p:420 (abstract).

Knopp MV, Strauss LG, Haberkorn UA, Blatter J, van Kaick G. Use of positron emission tomography for optimized therapy management of patients with lung tumors. *European Journal of Nuclear Medicine* 1990;16:560. (abstract)

Kubota K, Yamada S, Ishiwata K, Ito M, Ido T. Positron emission tomography for treatment evaluation and recurrence detection compared with CT in long-term follow-up cases of lung cancer. *Clinical Nuclear Medicine* 1992;17:877-81.

Madar I, Hoh C, Figlin RA, Holmes CE, Waters CE, Gambhir SS, et al. Cost effective staging of non-small cell lung carcinoma by whole body PET FDG imaging. *The Journal of Nuclear Medicine* 1995;36(5 Suppl.):57p-58p. (abstract)

Minn H, Zasadny KR, Quint LE, Wahl RL. Lung Cancer: Reproducibility of quantitative measurements for evaluating 2-[F-18]-fluoro-2-deoxy-d-glucose uptake at PET. *Radiology* 1995; 196:167-73.

Scott WJ, Gobar LS, Hauser LG, Sunderland JJ, Dewan NA, Sugimoto JT. Detection of scalene lymph node metastases from lung cancer. *Chest* 1995;107:1174-6.

Valk P, Pounds T, Hopkins D, Haseman M, Hofer G, Greiss H. Staging lung cancer by whole-body PET-FDG imaging. *The Journal of Nuclear Medicine* 1995;36(Suppl.):57p. (abstract)

Appendix 7

Systematic Review:
PET as a Diagnostic Test in Solitary Pulmonary Nodules

Author: Elizabeth Adams, R.R.T., M.P.H., *Management & Program Analyst, MDRC Technology Assessment Program*

Appendix 7

Systematic Review: PET as a Diagnostic Test in Solitary Pulmonary Nodules

The final literature database searches for the systematic reviews were performed on September 10, 1996; the assessment represents peer-reviewed literature published and indexed as of that date.

This Appendix to the PET assessment presents the results of the systematic review of PET in diagnosing solitary pulmonary nodules. A general rationale for the use of PET in oncology is supplied by Hawkins, et al., (1994) and Hoh, et al., (1994):

- many forms of cancer characteristically perturb tissue biochemical and physiological processes and PET imaging can be expected to detect the resulting abnormalities;
- reliance on tumor histology and anatomy limits the oncologist's tools for selecting optimal treatment;
- the ability to monitor metabolic responses to treatment could allow the early re-direction of therapy in patients who fail to respond to the first attempt at radiation or chemotherapy.

These and other authors (e.g., Price and Jones, 1995) report that PET studies in cancer are emerging as a major focus of the technology, both in basic research and in clinical investigations. Information gathered by the MDRC Technology Assessment Program from VA PET facilities corroborates that perception (see *Appendix 9: Experience With PET in VHA*).

Fluorine-18-fluorodeoxyglucose (FDG) is the most commonly employed radiopharmaceutical in PET cancer studies. Many neoplasms have high glycolytic rates, resulting in intracellularly trappedphosphorylated FDG that can be imaged with PET. Hawkins, et al. (1994), note that tumor-specific biochemical characteristics of glucose transport and phosphorylation may affect quantitative estimates of tumor glucose metabolism with FDG PET, and that investigations are under way to define these characteristics. However, these uncertainties may be of less concern with qualitative or semiquantitative FDG PET cancer studies because the primary intent of such studies is to detect and map tumor foci, not to rigorously quantify tumor glycolytic rates.

In some instances, PET imaging techniques have been modified to meet the needs of cancer diagnosis. Most PET systems allow axial fields of view (the length of the body encompassed by a series of cross sectional images) of approximately 10 cm. Cancer is frequently distributed beyond this field of view, and whole body image acquisition procedures have been developed (Hoh, et al., 1993). Since it is impractical to apply standard transmission scanning attenuation correction methods to these procedures, whole body PET imaging is primarily useful as a qualitative indicator of disease distribution.

Nieweg (1994) and Price and Jones (1995) define a number of potential applications for PET in oncology. These include:

- tumor detection (although PET images offer insufficient structural detail and should not be used to visualize anatomy; registration techniques to combine PET and anatomic imaging into a single image are under development to circumvent this limitation);
- staging (particularly using whole-body imaging methods) although there is a lower limit to the size of metastases that can be detected by PET;
- detection of local recurrence of disease, since anatomically-based imaging is often limited by the effects of treatment;
- prediction of tumor response to chemotherapy;
- treatment monitoring.

I. BACKGROUND

A. General sources

The discussion in this overview section, unless otherwise noted, is based on information provided by Lillington and Caskey (1993).

B. Description

This report will confine its discussion to those nodules that are solitary. A solitary pulmonary nodule (SPN) is a single spherical lesion within the lung not associated with hilar enlargement or atelectasis (incomplete expansion and/or collapse of lung tissue characterized on x-ray by local opacification), and whose size is generally less than 4.0 cm in diameter. Detection of multiple pulmonary nodules suggests a different group of diagnostic possibilities and a correspondingly different management approach.

C. Epidemiology

SPNs represent approximately 15% of all lung cancer diagnosed; in 1996, it is estimated that there will be 26,550 new cases of malignant SPNs in the United States (Cancer Facts & Figures 1996, American Cancer Society). The differential diagnoses of a SPN include many malignant and benign processes. Approximately 50% of SPNs are benign; infectious granulomas represent 80% of all benign SPNs and are caused predominately by coccidiomycosis, histoplasmosis, and tuberculosis (Midthun, 1993). Less common etiologies include hamartomas, noninfectious granulomas, infectious lesions, and vascular lesions.

The most common forms of malignant SPNs are bronchogenic carcinomas. According to the TNM staging system adopted by the American Joint Committee on End Results Reporting, a malignant SPN represents a clinical stage I lesion, which is potentially curable with resection (Mountain, 1993). Reported prevalence of malignant SPNs, ranging from less than 5% to greater than 70%, varies as a result of referral bias within each reported patient series. Lesions that have metastasized from extrathoracic tumors represent approximately 10-30% of all malignant SPNs (Midthun, 1993).

The following risk factors directly correlate with the probability of cancer in patients with a SPN: 1) patient's age; 2) patient's smoking history; 3) prior history of malignancy; 4) stability of lesion size on chest x-ray for 2 years; 5) presence of occult calcification within the nodule, and; 6) nodule size and characteristics of the nodule's edge as visualized on radiography. Additionally, the baseline prevalence of malignancy in the study population may suggest the likelihood of malignancy of the SPN. Exposure to benign diseases such as tuberculosis or a history of residence in areas endemic for coccidiomycosis or histoplasmosis will suggest a lesser likelihood of, but not rule out, malignancy.

D. Diagnosis

The vast majority of SPNs are incidental findings on a standard chest x-ray. Once the SPN is detected, the goal of clinical management is to choose the diagnostic approach most suited to the patient's clinical risk of malignancy, thus minimizing the number of thoracotomies for benign disease and expediting surgical resection of malignancies. Figure 1 represents a typical algorithmic approach to the clinical diagnosis and management of SPNs (Karlinsky, 1991).

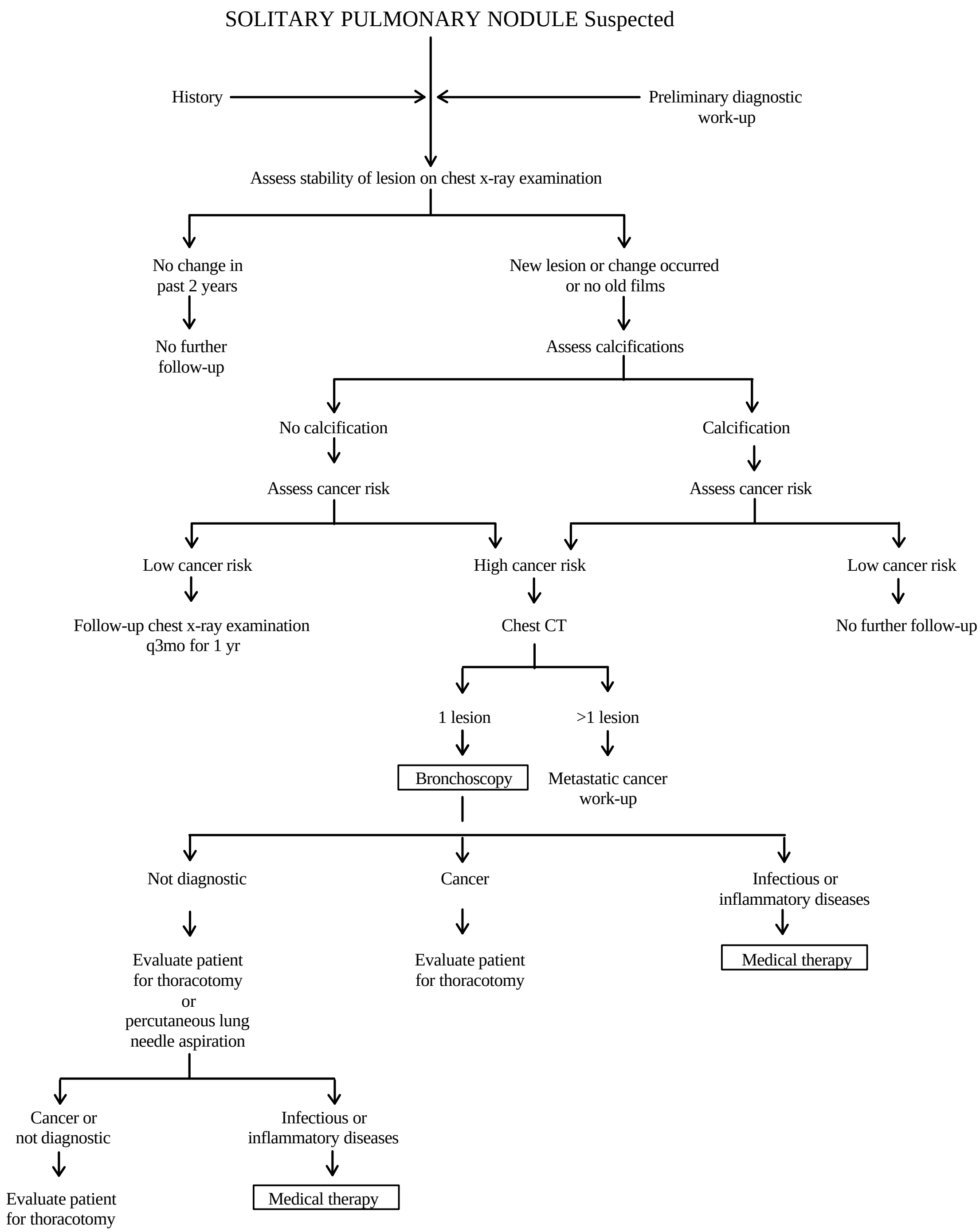


FIGURE 1. Algorithm for the management of solitary pulmonary nodules. (Karlinsky, 1991)

The patient's clinical risk of malignancy is estimated by weighing the patient's risk factors. These clinical data are collected through physical examination, review of the patient's medical history, and radiologic assessment. In a typically older veteran patient, who presents with an extensive smoking history and who is a surgical candidate, the probability of malignancy is often sufficiently high to proceed directly to thoracotomy without prior histologic confirmation. However, confirmation may be required if the patient is a high surgical risk or requests diagnostic confirmation of cancer prior to surgical resection. In the nonveteran population, where the probability of malignancy is likely to be lower, histologic confirmation may be needed to rule out benign disease and avoid an unnecessary thoracotomy.

Tissue may be sampled using several procedures including sputum collection, fiberoptic bronchoscopy, transthoracic needle aspiration biopsy (TTNA), thoracoscopy, and thoracotomy. Their diagnostic yields vary widely and depend on the procedures themselves, operator expertise, as well as size, type, and location of the nodule.

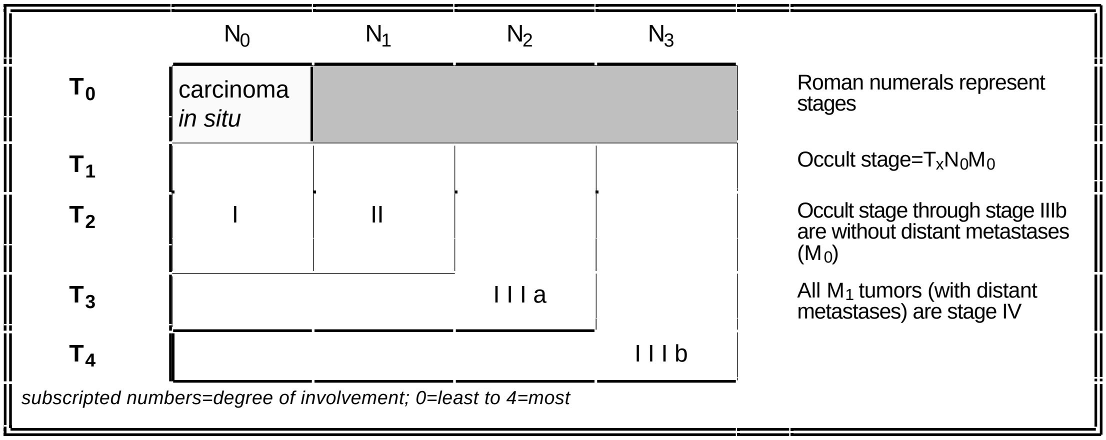
Sputum sampling has shown limited utility in patients with SPNs. Midthun (1993) reported ranges of diagnostic yield for bronchoscopy and TTNA of 20 to 80% and 43 to 97%, respectively. The lower yield of bronchoscopy may be attributed to the peripheral nature of the typical SPN. The wide range of diagnostic yield for TTNA may be attributed in part to the size of the nodule, with nodules larger than 2.0 cm in diameter associated with greater yield. Unlike bronchoscopy TTNA is a higher risk procedure with a reported incidence in pneumothorax of 15-30%, although only a small portion of these pneumothoraces may actually require treatment.

Minimally invasive surgical alternatives to thoracotomy for diagnosing indeterminate SPNs, such as thoracoscopy, are being studied and may contribute to the reduction in thoracotomies performed on patients with benign disease. Thoracotomy remains the definitive means for obtaining the diagnostic gold standard, but is the most invasive and exposes the patient to the risks of surgery.

E. Staging, treatment, and survival

Lung cancer staging assesses the extent of local and distant disease and involves two parts: 1) anatomic staging, and 2) physiologic staging or the ability of the patient to tolerate specific therapeutic interventions (performance status). Illustrated below is the TNM International Stage System (ISS) developed by the American Joint Committee on End Results Reporting used to describe the extent of primary tumor involvement (T stage), lymph node involvement (N stage) and distant metastasis (M stage), and to reflect prognosis and survival among homogenous patient groups with non-small cell lung carcinomas (Mountain, 1993).

Table 1 Lung Cancer TNM Staging System



Source: Mountain, 1993

Surgical removal of the malignancy and medical therapy for benign infectious or inflammatory diseases are the treatments of choice for diagnostically confirmed SPNs. For indeterminate SPNs the choices are to proceed with thoracotomy or to observe nodule change through serial chest x-rays (i.e., the "wait and watch strategy"). Observation is considered appropriate for patients with a very low probability of malignancy, although the potential adverse affect of the delay in resection on patient survival is controversial.

F. Potential roles for PET

Conventional wisdom supports a prevalence of malignancy of 30-50% among resected indeterminate SPNs based primarily on a report from the U.S. Veterans Administration Armed Forces Cooperative Study (Steele, 1963). Accordingly, there is a need to improve diagnostic accuracy, with the hope of identifying preoperatively a larger number of benign lesions and avoiding unnecessary thoracotomies. The standard radiologic method of choice for evaluating SPNs is computerized tomography (CT) because of its enhanced visibility and morphologic detail.

CT is used in many capacities in the evaluation of SPNs: 1) to determine the number of nodules; 2) to assess nodule size; 3) to determine shape and characteristics of the nodule's edge; 4) to visualize evidence of calcification and; 5) to serve as a guide for biopsy procedures. Iodinated contrast material and high resolution CT densitometry (HRCT) are used to enhance conventional CT. Preliminary studies describing improved detection of malignant SPNs with CT enhanced with iodinated contrast material have been reported (Swensen, 1995).

HRCT employs a reference "phantom" to indirectly demonstrate "occult" calcification, the presence of which shows a strong, but not definitive, likelihood of benignity in approximately 50% of nodules that appear noncalcified by standard imaging techniques. Although HRCT provides exceptional morphologic detail, limitations in its ability to differentiate benign from malignant lesions have been reported. These limitations become more apparent with decreasing nodule size.

MRI has been proposed as a possible adjunct to CT in the clinical work up, but has not demonstrated greater benefit over CT. Consequently, many lesions classified as indeterminate before CT are still indeterminate afterward and require evaluation using biopsy procedures most appropriately matched to the patient's level of clinical risk.

PET with FDG has recently been proposed as a potential solution for improving the noninvasive determination of benign from malignant SPNs. Current literature suggests that a PET scan would likely follow conventional imaging, specifically CT, in the diagnostic work up. The utility of PET in differentiating benign from malignant pulmonary nodules less than 3 cm, and occasionally, 4 cm in diameter (i.e., those nodules likely to be indeterminate) is being assessed, because nodules greater than 3 cm in diameter have a higher probability of malignancy.

II. RESULTS

Ten articles were selected from the MEDLINE and other database searches and from the bibliographies of initially retrieved articles as meeting the screening criteria. After review, 6 (60%) were found to meet the inclusion criteria for assignment to the following levels of the diagnostic efficacy hierarchy (Fryback and Thornbury, 1991; *Appendix 2: Assessing Diagnostic Technologies*): 2 met the definition of technical efficacy (see Reference List; full data abstraction tables for Technocal Efficacy studies are on file with the MDRC Technology Assessment Program); 2 met the criteria of diagnostic accuracy efficacy (Table 4); 2 met the criteria for diagnostic accuracy efficacy and diagnostic thinking efficacy (Table 5). The MDRC Technology Assessment Program was unable to locate any studies which addressed the impact of PET in the clinical management of these patients or on treatment outcomes.

Two studies not included in Tables 4 or 5 provided technical efficacy data that may provide useful information for subsequent diagnostic efficacy studies. Lowe, et al., (1995) assessed the optimal time for emission data acquisition using dynamic PET imaging. Duhaylongsod, et al., (1995a) assessed retrospectively the relationship between glucose metabolism measured by PET and tumor doubling time on radiography.

All studies presented in Tables 4 and 5 reporting diagnostic accuracy for PET in evaluation of SPNs are derived from case series, providing Level V (i.e., the weakest) evidence of any association between the use of a technology and improved patient outcomes. Operating characteristics from these series are based on a disproportionate number of cases to internal controls. The inclusion criteria varied across series with respect to nodule size and definition of indeterminate focal lesions, which may have included ill-defined infiltrates and pulmonary masses. Since there was either a strong or definite likelihood of work-up bias in these studies, none met the strict evidence-based medicine criteria for blinding. However, with the exception of Gupta, et al., (1996) all studies provided some information on blinding of their test readers to the biopsy gold standard.

A. Characterizing indeterminate solitary pulmonary nodules

Two studies presented in Table 4 evaluated the diagnostic accuracy of PET in the work up of SPNs. Dewan, et al., (1995) reported results comparing PET and transthoracic fine needle aspiration biopsy (TTNA) in patients who had undergone both procedures to diagnose peripheral SPNs. Patients with lung masses > 3 cm, hilar lesions, and multiple pulmonary nodules were also included, implying a high index of suspicion for malignancy. That the decision to perform TTNA may have been influenced by the PET results, implying a strong association between the test result and determination of the final diagnosis, was unclear. These authors also reported a significantly higher rate of complications (incidence of pneumothorax and of chest tubes) from TTNA than from PET.

Bury, et al., (1996) evaluated the complementary role of PET in characterizing indeterminate SPNs at a point in the work up after radiography. Nodules were noncalcified and ranged from 0.5 cm-4.5 cm in size. Patient selection was limited to those scheduled for biopsy determination, suggesting a high index of suspicion of malignancy. Results from both of these studies should be viewed cautiously, as the degree of bias is significant.

Table 5 presents data abstracted from two studies that assessed the quantitative importance of PET in the diagnostic work up (diagnostic thinking efficacy) of SPNs and diagnostic accuracy. Duhaylongsod, et al., (1995b) reported operating characteristics and likelihood ratios based on a quantitative methodology for patients with an indeterminate nodule ≤ 4 cm in diameter. The cut-off value of ≥ 2.5 to define malignancy was determined retrospectively from a receiver operating characteristic (ROC) curve. A subgroup analysis for lesions < 3 cm in diameter yielded similar results to those categorized as < 4 cm in diameter.

Likelihood ratios for quantitative standard uptake ratio (SUR) values, which have a continuous distribution, were reported in an attempt to determine the incremental value of PET in the work up of SPNs, and may provide additional information on the relationship between the SUR and severity of disease. SUR values located farther from the cut-off may assist clinical decision making with greater certainty than those values located closer to the cut-off. Diagnostic certainty and subsequent treatment decisions may also vary with the choice of cut-off. The degree to which PET would have changed diagnostic certainty and treatment decisions, particularly with respect to the number of thoracotomies spared, was not systematically assessed in this study, and firm conclusions on the incremental value of PET in this diagnostic process can not be drawn.

These authors presented a hypothetical cost analysis to assess the economic impact of PET in this patient population. Thirteen patients, who were initially excluded from the study because of insufficient follow-up and diagnostic confirmation, were included in these calculations. These authors calculated a reduction in overall costs of $397,062 when using PET in the work-up. The MDRC Technology Assessment Program revised the potential cost savings to be $158,934 to reflect only the results of the 87 patients initially included in the study. However, both of these calculations were based on a narrow group of assumptions using hospital charge data that may not accurately reflect true costs and may not be sufficiently generalizable to other patient populations. Variations in case-mix and optimal cut-off value, which may occur across populations, and the hypothetical nature of the cost analysis suggest that these results are preliminary.

Gupta, et al., (1996) assessed the diagnostic accuracy of qualitative PET scans in patients with indeterminate SPNs ≤ 3 cm in size and compared methods for computing the probability of malignancy based on PET results, patient's age, and nodule size. This study may not have met evidence-based medicine criteria, because blinding of the PET readers to the diagnostic gold standard was not noted. Limitations in reporting with respect to blinding of the PET readers to other clinical and radiographic data prevented determination of the conditional independence between tests in a sequence, an assumption required of Bayes' Theorem in sequential testing, and the subsequent incremental value of PET in the work up.

In their calculations the authors used a pre-test probability of malignancy of 0.40 based on the overall prevalence of malignancy in the general population. In patients (particularly veterans), who would be referred to a PET center at a point in the diagnostic process after radiologic imaging but before tissue sampling, the prevalence of malignancy is likely to be much higher. This study may not provide valid estimates of the quantitative importance of PET in these patients and should be interpreted cautiously.

III. SUMMARY

Table 3 lists preliminary studies of the accuracy of PET in diagnosing indeterminate solitary pulmonary nodules (i.e., those lesions with equivocal findings on CT) at a point in the diagnostic process after CT but before biopsy procedures. All four studies of diagnostic accuracy were case series with a high proportion of malignant cases reflecting both the study inclusion criteria and, indirectly, the relative accuracy of current modalities used prior to PET in the diagnostic work up to identify benign disease.

Each study varied in its inclusion criteria with respect to maximal lesion size and image characteristics (pulmonary mass, ill-defined infiltrates, focal lesions). No data comparing PET with alternative imaging technologies such as CT were presented, although one study (Dewan, 1995) attempted to assess the complementary role of PET with an invasive needle biopsy procedure (TTNA) to characterize SPNs. Two of these studies (Duhaylongsod, 1995b and Gupta, 1996) also attempted to quantify the importance of PET in the diagnostic work up of SPNs. Limitations in reporting and study design preclude drawing firm conclusions from these series.

IV. DISCUSSION

The attempt by some of these studies to characterize comparison groups quantitatively warrants further discussion. Consideration of the cut-off point used in quantitative analysis will depend on the consequences of limiting false negative results and of accepting false positive results. This value will also be influenced by the pre-test probability of disease of the study population and by the heterogeneity of normal, benign, and malignant lung tissue. In a recently published case control study Miyauchi, et al., (1996) demonstrated the effect of regional variations of FDG uptake within normal lungs on the range of reported results, particularly with respect to small lung nodules found in lower lung fields. Variations in normalization procedures used in semi-quantitative analyses may further influence the choice of cut-off (Schomburg, 1996). Determination of the clinical efficacy of PET using quantitative methods requires defining the optimal cut-off point from a much larger and broader study population, and subsequently applying it to studies designed to determine diagnostic accuracy.

Methodologies have been developed to enhance the interpretation of diagnostic information in patients with SPNs. They include the application of probabilistic reasoning methods such as Bayesian analysis and decision analysis, and computer-assisted analytic techniques using neural networks, thresholding, or profile matching.

Bayesian analysis combines radiographic findings, such as location, size, and edge characterization on CT and/or x-ray, with clinical information, such as age and smoking history, to estimate the probability that a nodule in an individual patient is malignant. An estimation of likelihood ratios for various individual radiographic and clinical characteristics on previously evaluated patients with SPNs must first be determined from the literature. The case mix of the sample population will likely affect the derivation of these ratios (Gurney, Part 1., 1993).

The utility of this analysis may be found in its examination of all of the pertinent clinical and radiologic information, rather than reliance on the results of one test. For example, Gurney and associates focused on interpretation of detected nodules, rather than detection itself, by comparing the accuracy of diagnosis of SPNs using Bayesian analysis with the accuracy of independent estimation by expert radiologists (Gurney, Part 2., 1993). This study found that the readers using Bayesian analysis performed better than the expert readers (i.e., fewer misclassifications of the nodules) in reading both individual studies and patients' combined studies. There was better concordance among readers who used Bayesian analysis. Moreover, additional clinical information did not necessarily improve the readers' performance in either group.

Decision analysis has been used to assess the relative value of clinical strategies in the presence of an uncertain diagnosis (Lillington, 1989). The probability of cancer (pCA) in any given case will depend on the clinical factors present in that case. Variations in the magnitude of the calculated values for pCA will affect the utility of each strategy. A high pCA suggests the use of thoracotomy, whereas a low pCA supports the observation strategy. A pCA in the middle range supports the use of advanced needle biopsy procedures followed by thoracotomy, if needed. Lillington (1989) found that patient preferences should be considered when the difference in expected utility between biopsy and surgery tends to be small. With decision analysis, it is also possible to approximate the value for the deleterious effect of delaying the resection of a malignant nodule.

Preliminary work has been performed using digital techniques to assist the radiologist in the diagnosis of SPNs (Lo, 1993 and Gurney, 1995). The specific aim of these techniques is to enhance true positive detection, thereby reducing the number of false positive results, through the use of computer image processing. The results of these studies have not shown greater benefit over conventional methods to date, although their application will likely increase with the increasing use of digital imaging technology.

V. SUGGESTIONS FOR FURTHER PET RESEARCH

The potential benefits of PET in the diagnosis and management of solitary pulmonary nodules (SPNs) have been purported. However, no evidence published to date definitively supports the routine use of PET in these patients. Four studies attempted to define the operating characteristics of PET as a diagnostic test in this area, including two that also attempted to quantify the importance of PET in the diagnostic work up of SPNs. All have methodologic shortcomings, and their results should be interpreted cautiously.

CT has a more established role in the clinical management of solitary pulmonary nodules, is more widely available with associated lower costs, and provides valuable anatomical detail not always available with PET. Conventional wisdom has defined the limitations of CT with respect to characteristics of resected indeterminate nodules based primarily on data that were reported from surgical series comprising a significant proportion of young patients with shorter smoking histories, and that were derived from studies conducted prior to the use of CT or during its early stages of diffusion.

A recent study conducted by Rubins and Rubins (1996) reported an increasing proportion of malignancy in resected indeterminate SPNs over the last fourteen years (from 55% in 1981 to 60% in 1983 and from 90% in 1990 to 100% in 1994) at a single university-affiliated VA Medical Center. They attributed these trends to improvements in the ability to diagnose benign SPNs preoperatively, primarily through the use of CT. In the presence of these trends, a technology such as PET would need to demonstrate significant improvements in patient outcome or reductions in associated costs in order to justify its role in the clinical work up of SPNs.

Less resource-intensive analytical models exist to provide the framework with which to assess the impact of diagnostic imaging in the management of SPNs. Nevertheless, these models require that evidence of both operating characteristics and underlying characteristics of the study population exist prior to implementation.

Contributions from other investigators working with larger and well-defined patient populations and comparing PET to existing modalities will be needed to refine the characteristics of PET as a diagnostic tool in SPNs, specifically as they pertain to the veteran population, and to establish a base for further research. Any attempt to expand the role of PET into earlier stages in the diagnostic work up would require an evaluation designed accordingly.

In this context, future research within VA should focus on:

1) establishment of a PET registry, which would provide a range of data on demographic and clinical characteristics of patients in whom PET studies are performed, and on their clinical outcomes in a variety of settings;

2) establishment of estimates of a cut-off point to define disease and of subsequent diagnostic accuracy;

3) studies designed to assess the role and impact of PET in the diagnostic work-up of SPNs (eg., to avoid unnecessary surgery, to replace needle biopsy, or to replace conventional imaging in detecting disease)

Table 2 Summary of the Literature: Diagnostic accuracy efficacy studies of PET in solitary pulmonary nodules

Notes: All of the studies in the table are case series (Level V evidence) and met most of the evidence-based medicine criteria for diagnostic test evaluations. None of the studies met strict evidence-based medicine criteria for blinding, but all except Gupta, et al., (1996) provided information on the comprehensiveness of blinding of test interpreters to the biopsy gold standard. Blinding of PET interpreters to other clinical and radiologic data varied across studies.

Internal controls (i.e., those with benign masses) were used in each study, and it was possible to calculate sensitivity and specificity for PET in those studies. The pre-test probability of disease in these study populations was very high, and predictive values were not reported. Each study varied in inclusion criteria with respect to maximal lesion size and image characteristics (pulmonary masses, ill-defined infiltrates, focal lesions). Operating characteristics from these studies should be interpreted with caution.

Where substantial duplication in purpose of study, patients studied, and results in multiple studies from the same institution could be inferred, only the most recent, largest, most rigorously designed, or most comprehensive was included in the table. While data from Dewan, et al., (1995) and Gupta, et al., (1996) are likely derived from the same patient population, these studies addressed different purposes, and inclusion of both was felt to be warranted. Studies reviewed but not included are listed under "References".

Abbreviations are listed at the end of the table.

Role	***Study***	***N***	***Operating Characteristics****			***Evidence-Based Medicine Criteria*****			***Methodologic Quality Grade*****
			PET	***TTNA***	***other***	***comparison group***	***histologic gold standard***	***blinding***	
Defining unknown SPN	Dewan, et al., 1995	26 malignant lesions 9 benign lesions	Se=100% Sp=78% accuracy=94%	Se=81% Sp=100% accuracy=86%		+ internal	+	partial	**D**
	Bury, et al., 1996	33 malignant cases 17 benign cases	Se=100% Sp=88%			+ internal	+	+	**C**
	Duhaylongsod, et al., 1995b	59 malignant cases 28 benign cases	*for lesions < 4 cm* Se=97% Sp=81% accuracy=92%			+ internal	+	+	**C**
	Gupta, et al., 1996	45 malignant cases 16 benign cases	Se=93% Sp=88% accuracy=92%			+ internal	+	unclear	**C**

Se, sensitivity
Sp, specificity
TTNA, transthoracic needle aspiration biopsy

** operating characteristics defined in Appendix 2: Assessing Diagnostic Technologies, pages 5-7*
*** Appendix 2, page 8*
**** Appendix 2, page 9*

Table 3 Diagnostic efficacy of FDG PET in diagnosing solitary pulmonary nodules

Notes: The studies in this table are case series and met all or most of the evidence-based criteria for diagnostic test evaluations. Internal controls (i.e. those with benign masses) were used, and it was possible to calculate sensitivity and specificity for PET. There was a high ratio of malignant to benign cases; therefore, predictive values were not reported. Each study varied in inclusion criteria with respect to maximal lesion size and image characteristics (pulmonary masses, ill-defined infiltrates, focal lesions). Blinding of PET interpreters to other clinical or radiologic findings was not explicit, and the incremental value of PET could not be determined. Operating characteristics should be interpreted cautiously.

Where substantial duplication in purpose of study, patients studied, and results in multiple studies from the same institution could be inferred, only the most recent, largest, most rigorously designed, or most comprehensive was included in the table. Studies reviewed but not included are listed under "References".

Abbreviations are listed at the end of this table.

Study	*Patients/Methods*	*Results/Comments*
Dewan, et al., 1995 (*Creighton University and Veterans Affairs Medical Center, Omaha, Nebraska)*	**Purpose** a retrospective analysis to compare PET to transthoracic fine-needle aspiration biopsy (TTNA) in diagnosing peripheral solitary pulmonary lesions **Cases** 33 patients with 35 lung lesions who had undergone both PET and TTNA (26 malignant, 9 benign) • 22 SPNs (< 3 cm); 4 hilar lesions; 8 lung masses (> 3 cm); 1 with multiple pulmonary nodules **Methods** • all patients had chest x-ray and CT • decision to perform TTNA made by primary physician aware of PET results • qualitative PET performed • both PET interpreters blinded to biopsy • TTNA performed under CT guidance • PET and TTNA compared to biopsy results **Limitations of study design** • retrospective analysis • number of cases and internal controls not equivalent • blinding of PET interpreters to clinical or radiologic data not noted • test result and determination of final diagnosis not independent	**Defining SPN lesion (26 malignant lesions, 9 benign lesions)** • PET: Se=100%; Sp=78%; accuracy=94% • TTNA: Se=81%; Sp=100%; accuracy=86% • no statistically significant differences reported between two techniques **Complications** • pneumothorax: PET=0/35 (0%) TTNA=16/35 (46%) p=0.0001 • chest tube: PET=0/35 (0%) TTNA=9/35 (26%) p=0.0039 **Discussion** • authors report study size limitation; a highly select group may affect generalizability of results • authors report interobserver agreement, but not measured

Study	Patients/Methods	Results/Comments
Bury, et al., 1996 *(CHU, Liège, Belguim)*	**Purpose** to prospectively assess the diagnostic accuracy of FDG PET in diagnosing solitary pulmonary nodules (SPN) **Cases** 50 patients with indeterminate SPNs after chest x-ray and CT, ranging from 0.5 cm to 4.5 cm in size (33 malignant, 17 benign) **Methods** • all patients fasted for six hours prior to PET • PET interpreted visually and classified as no uptake, moderate or intense • independent interpretation by two groups of nuclear medicine physicians reached by consensus and who had knowledge of chest x-ray but not CT, and blinded to biopsy results • PET results compared to biopsy results **Limitations in study design** • number of cases and internal controls not equivalent (high prevalence of malignancy) • patient selection limited to those scheduled for invasive procedure (biased toward those with a high index of suspicion for malignancy) • partial blinding of PET readers to other clinical data • independence of test result and determination of disease unclear	**Defining SPN lesion (33 malignant cases, 17 benign cases)** • PET: Se=100%; Sp=88% • CT: no data reported • two false positive results due to tuberculosis and chronic nonspecific inflammation • mean size (range) of nodules: malignant=3 cm (1.5 cm- 4.5 cm) benign= 1.8 cm (0.5 cm- 3.5 cm) **Authors' comments** no difference in FDG uptake across histopathologic types was observed; quantitative analysis may be needed for clarification

Abbreviations: Se, sensitivity
Sp, specificity
CT, computerized tomography

*indicated calculated by MDRC TA Program from data supplied in published article

Table 4 Diagnostic thinking efficacy of FDG PET in solitary pulmonary nodules

Notes: Both studies are case series (Level V evidence) with internal controls (i.e. those with benign masses), and it was possible to calculate sensitivity and specificity for PET. All patients in these studies had suspected or biopsy-proven lung cancer (i.e., the pre-test probability of disease in the study populations was very high); therefore, predictive values were not reported. The study by Gupta, et al., (1996) may not have met the evidence-based medicine criteria for blinding. Each study varied in inclusion criteria with respect to maximal lesion size and image characteristics (pulmonary masses, ill-defined infiltrates, focal lesions). Operating characteristics and likelihood ratios should be interpreted with caution.

Where substantial duplication in purpose of study, patients studied, and results in multiple studies from the same institution could be inferred, only the most recent, largest, most rigorously designed, or most comprehensive was included in the table. Studies reviewed but not included are listed under "References".

Abbreviations are listed at the end of this table.

Study	*Patients/Methods*	*Results/Comments*
Duhaylongsod, et al., 1995b (*Duke University, Durham North Carolina)*	**Purposes** to evaluate the diagnostic accuracy of FDG PET in differentiating benign from malignant focal pulmonary lesions, suspected primary lesions or recurrent cancer **Cases** 87 patients with indeterminate focal pulmonary lesions by chest x-ray and CT *(13 patients had been excluded lacking firm pathologic diagnosis and < 2 years follow up)* • 59 malignant; 28 benign • 79 SPNs (defined as focal abnormalities ≤ 4 cm diameter); 11 pulmonary masses; 10 ill-defined infiltrates • included 16 patients evaluated for recurrent disease **Methods** • two PET cameras used in study • ROIs chosen; SURs calculated by one nuclear medicine physician blinded to patient history, physical exam, and labs, including biopsy results • biopsy (n=84) and follow up > 2 years (n=3) confirmation obtained • SURs compared; mean ± SD reported • Se, Sp, accuracy, and likelihood ratios calculated • cost analysis performed to assess economic impact of two strategies, immediate thoracotomy and PET, on diagnosis and management of focal pulmonary lesions with following assumptions, using hospital charge data: - SUR ≥ 2.5 lead to thoracotomy, SUR < 2.5 warranted observation - all indeterminate SPNs < 3 cm diameter - pretest probability of malignancy=50% - PET Se=97%; Sp=82% - thoracotomy complication rate=zero - total hospital stay for thoracotomy=5 days **Limitations of study design** • choice of SUR cutoff determined retrospectively • number of cases and internal controls not equivalent (high pre-test probability of disease) • study included subjects with disease other than SPNs • independence of test result and determination of final diagnosis unclear	**Defining unknown focal disease (59 malignant cases, 28 benign cases)** overall SUR: malignant=6.6 ± 3.1 vs. benign= 2.0 ± 1.6 (*p* = 0.0001) *based on cut-off SUR ≥ 2.5 chosen for malignancy determined from ROC curve* Se=97%; Sp=82%; accuracy=92% **Defining unknown SPNs (45 carcinomas, 22 benign)** SUR: malignant=5.5 ± 2.1 vs. benign= 1.7 ± 1.1 **Defining unknown pulmonary masses (10 malignant, 1 benign)** SUR: malignant=8.7 ± 3.8 vs. benign= 1.3 **Defining unknown pulmonary infiltrates (4 malignant, 5 benign)** SUR: malignant=5.1 ± 2.0 vs. benign= 2.8 ± 2.1 **Defining unknown SPNs < 3 cm diameter (31 malignant, 16 benign)** *based on cut-off SUR ≥ 2.5 chosen for malignancy determined from ROC curve* PET: Se=100%; Sp=81%; accuracy=94% **Likelihood ratios (LR) for five levels of SUR computed from FDG PET (malignant/benign cases)** *Se and Sp data used for calculations was not noted* SUR ≥ 6.0 LR= 16.136 (34/1) SUR 4.0-5.9 LR:= 3.085 (13/2) SUR 2.5-3.9 LR:= 1.582 (10/3) SUR 1.5-2.49 LR= 0.095 (2/10) SUR < 1.5 LR:= 0.000 (0/12) **Cost analysis** • strategy using PET resulted in 41 fewer nontherapeutic operations and reduced overall costs by 24.8% (Δ= $397, 062) based on all 87 patients + 13 patsxluded from the study • cost savings calculated by MDRC Technology Assessment Program based on 87 patients included in the study= $158,934 • conservative cost estimates did not account for extended length of stay, intensive care management, other discomfort, lost wages, or expenses from other procedures **Other findings** • false positives attributed to active infections • one false negative attributed to small lesion size (4 mm) • in patients evaluated for recurrent disease, all benign cases (n=10) had SURs < 2.5 • authors stress need to establish cost-effectiveness before widespread application

Study	Patients/Methods	Results/Comments
Gupta, et al., 1996 *(West Virginia University, Morgantown, West Virginia) (data collected at Creighton University and Omaha VAMC, Nebraska)*	**Purpose** • to assess the diagnostic accuracy of PET in the evaluation of solitary pulmonary nodules (SPNs) • to compare methods for computing the probability of malignancy in SPNs based on PET versus several risk factors **Cases** 61 patients with indeterminate nodules (0.6 cm-3 cm in size) based on chest x-ray and CT (45 malignant cases, 16 benign cases) **Methods** • all patients had chest x-ray and CT interpreted independently prior to PET • patients fasted for 4 hours before PET scans • PET analyzed qualitatively by two observers • DURs calculated for semiquantitative analysis • for patients in whom no nodule could be detected on PET, ROI was extrapolated from radiographic imaging • PET results compared to histology; one patient followed-up for 2 years **LImitations of study design** • blinding of PET to other clinical or radiographic data, or to biopsy results, not reported • number of cases and internal controls not equivalent (high pre-test probability of disease) • pre-test probability of disease used by authors (0.40) did not take into account the clinical data obtained prior to PET	**Defining unknown SPN (45 malignant cases, 16 benign cases)** PET: Se=93%; Sp=88%; accuracy=92% • 3 false negatives due to adenocarcinoma; 2 false-positive findings due to granuloma with histoplasmosis **Detecting hilar/mediastinal lymph adenopathy** 12 patients had confirmed hilar/mediastinal lymphadenopathy; in 5 patients nodal involvement was not suspected prior to PET, but PET accurately identified all abnormalities **Likelihood ratios based on PET results (45 malignant cases, 16 benign cases)** *assuming pre-test probability of disease=0.40; Se=93%; Sp=88%* LR=7.464; probability of malignant nodule, given a positive PET scan=0.833 LR=0.075; probability of malignant nodule, given a negative PET scan=0.047 **Likelihood ratios based on age** LR=0.405; probability of malignant nodule, given age < 60 years=0.213 LR=0.915; probability of malignant nodule, given age between 60-69 years=0.380 LR=3.376; probability of malignant nodule, given age between 70-89 years=0.693 **Likelihood ratios based on nodule size** LR=0.400; probability of malignant nodule, given nodule size ≤ 1.0 cm=0.211 LR=0.828; probability of malignant nodule, given nodule size between 1.1-1.9 cm=0.356 LR=2.064; probability of malignant nodule, given nodule size ≥ 2.0 cm=0.580 **Authors' comments** • authors reported interobserver variability < 5%, but no supporting data presented • all benign nodules < 2.5 cm in size; 14/16 < 2 cm in size • 11/45 malignant nodules < 2 cm in size • both nodules < 1 cm in size were accurately detected with PET • no correlation between DUR indices and histologic type • the probability of malignancy increases with nodule size, patient's age, and a positive PET scan • simultaneous pre-operative staging for hilar/mediastinal lymph nodes is an additional advantage of PET in patients with malignancies

Abbreviations:
CT, computerized tomography
ROC, receiver operating characteristic
ROI, region of interest
DUR, differential uptake ratio
SUR, standard uptake ratio
Se, sensitivity
Sp, specificity

VI. REFERENCES Background and studies meeting evidence-based medicine criteria for evaluations of diagnostic tests

American Cancer Society. *Cancer Facts & Figures-1996.* New York: National Media Office-ACS; 1996.

Bury T, Dowlati A, Paulus P, Corhay JL, Benoit T, Kayembe JM, *et al.* Evaluation of the solitary pulmonary nodule by positron emission tomography imaging. *European Respiratory Journal.* 1996;9:410-14.

Cummings SR, Lillington GA, Richard RJ. Estimating the probability of malignancy in solitary pulmonary nodules: a Bayesian approach. *American Review of Respiratory Disease*. 1986; 134:449-52.

Dewan NA, Reeb SD, Gupta NC, Gobar LS, Scott WJ. PET-FDG imaging and transthoracic needle lung aspiration biopsy in evaluation of pulmonary lesions: a comparative risk-benefit analysis. *Chest.* 1995;108:441-6.

Duhaylongsod FG, Lowe VJ, Patz EF, Vaughn AL, Coleman RE, Wolfe WG. Detection of primary and recurrent lung cancer by means of f-18 fluorodeoxyglucose positron emission tomography (FDG PET). *The Journal of Thoracic and Cardiovascular Surgery.* 1995;110:130-140. (b)

Gupta NC, Maloof J, Gunel E. Probability of malignancy in solitary pulmonary nodules using fluorine-18-FDG and PET. *The Journal of Nuclear Medicine*. 1996;37:943-8.

Gurney JW. Determining the likelihood of malignancy in solitary pulmonary nodules with Bayesian analysis: Part 1. Theory. *Radiology.* 1993;186:405-13.

Gurney JW, Lyddon DM, McKay JA. Determining the likelihood of malignancy in solitary pulmonary nodules with Bayesian analysis: Part 2. Application. *Radiology.* 1993;186:415-22.

Gurney JW, Swensen SJ. Solitary pulmonary nodules: determining the likelihood of malignancy with neural network analysis. *Radiology.* 1995;196:823-9.

Harvey JC, Beattie EJ. Surgical treatment of solitary and multiple metastatic tumors to the lung. *Comprehensive Therapy* 1993;19(5):238-41.

Henschke CI, Miettinen OS, Yankelevitz DF, Libby DM, Smith JP. Radiographic screening for cancer: proposed paradigm for requisite research. *Clinical Imaging.* 1994;18:16-20.

Karlinsky JB, Lau J, Goldstein RH. Solitary pulmonary nodule. In: Decker BC, ed. *Decision Making in Pulmonary Medicine.* Philadelphia: Mosby Year Book, Inc.; 1991.

Khan A, Herman PG, Vorwerk P, Stevens P, Rojas KA, Graver M. Solitary pulmonary nodules: comparison of classification with standard, thin-section, and reference phantom CT. *Radiology.* 1991;179:477-81.

Lillington GA, Caskey CI. Evaluation and management of solitary and multiple pulmonary nodules. *Clinics in Chest Medicine.* 1993;14(1):111-9.

Lillington GA, Cummings SR. Decision analysis approaches in solitary pulmonary nodules. *Seminars in Respiratory Medicine.* 1989;10(3):227-31.

Lillington GA. Management of solitary pulmonary nodules. *Disease A Month.* 1991;37:271-318.

Lo S-CB, Freedman MT, Lin J-S, Mun SK. Automatic lung nodule detection using profile matching and back-propagation neural network techniques. *Journal of Digital Imaging.* 1993;6(1):48-54.

Midthun DE, Swensen SJ, Jett JR. Approach to the solitary pulmonary nodule. *Mayo Clinic Proceedings.* 1993;68:378-85.

Mitruka S, Landreneau RJ, Mack MJ, Fetterman LS, Gammie J, Bartley S, *et al.* Diagnosing the indeterminate pulmonary nodule: percutaneous biopsy versus thoracoscopy. *Surgery*. 1995; 118:676-84.

Mountain, C. Lung cancer staging classification. *Clinics in Chest Medicine.* 1993;14(1):43-51.

Rubins JB, Rubins HB. Temporal trends in the prevalence of malignancy in resected solitary pulmonary lesions. *Chest.* 1996;109:100-3.

Steele JD. The solitary pulmonary nodule: report of a cooperative study of resected asymptomatic solitary pulmonary nodules in males. *Journal of Thoracic and Cardiovascular Surgery.* 1963;46(1):21-39.

Swensen SJ, Brown LR, Colby TV, Weaver AL. Pulmonary nodules: CT evaluation of enhancement with iodinated contrast material. *Radiology.* 1995;194:393-8.

Swensen S, Harms GF, Morin RL, Myers JL. CT evaluation of solitary pulmonary nodules: value of 185-H reference phantom. *AJR American Journal of Roentgenology.* 1991;156:925-9.

Takanashi N, Nobe Y, Asoh H, Yano T, Ichinose Y. The diagnostic accuracy of a solitary pulmonary nodule, using thin-section high resolution CT: a solitary pulmonary nodule by HRCT. *Lung Cancer.* 1995;13:105-12.

Zwirewich CV, Vedal S, Miller RR, Müller NL. Solitary pulmonary nodule: high-resolution CT and radiologic-pathologic correlation. *Radiology.* 1991;179:469-76.

VII. REFERENCES: Technical efficacy studies (not included in the tables)

Duhaylongsod FG, Lowe VJ, Patz EF, Vaughn AL, Coleman RE, Wolfe WG. Lung tumor growth correlates with glucose metabolism measured by fluoride-18 fluorodeoxyglucose positron emission tomography. *Annals of Thoracic Surgery.* 1995;60:1348-52. (a)

Lowe VJ, DeLong DM, Hoffman JM, Coleman RE. Optimum scanning protocol for FDG-PET evaluation of pulmonary malignancy. *The Journal of Nuclear Medicine.* 1995;36:883-7.

VIII. REFERENCES: Excluded studies

Exclusion criteria included:

- number of cases < 12
- duplicated or superseded by subsequent or concurrent study from the same institution
- insufficient information to judge comparability of case and control groups, details of imaging protocol, whether visual or quantitative analysis of PET data used, or type of PET data analysis used
- abstract, not peer reviewed

Dewan NA, Gupta NC, Redepenning LS, Phalen JJ, Frick MP. Diagnostic efficacy of PET-FDG imaging in solitary pulmonary nodules: potential role in evaluation and management. *Chest.* 1993; 104(4):997-1002.

Gupta NC, Frank AR, Dewan NA, Redepenning LS, Rothberg ML, Mailliard JA, *et al.* Solitary pulmonary nodules: detection of malignancy with PET with 2-[f-18]-fluoro-2-deoxy-d-glucose. *Radiology.* 1992;184:441-4.

Gupta NC, Dewan NA, Phalen JJ, Rothberg ML, Shiue CY, Frick MP, *et al.* Diagnostic efficacy of PET-FDG imaging in the differential diagnosis of a solitary pulmonary nodule. *Radiology.* 1991;181:152 (abstract)

Patz EF, Lowe VJ, Hoffman JM, Paine SS, Burrowes P, Coleman RE, Goodman PC. Focal pulmonary abnormalities: evaluation with f-18 fluorodeoxyglucose PET scanning. *Radiology.* 1993;188:487-90.

Appendix 8

Systematic Review:
PET as a Diagnostic Test in Alzheimer's Disease

Author: Karen Flynn, D.D.S., M.S., *Manager, MDRC Technology Assessment Program*

Appendix 8

Systematic Review: PET as a Diagnostic Test in Alzheimer's Disease

The final literature database searches for the systematic reviews were performed on September 10, 1996; the assessment represents peer-reviewed literature published and indexed as of that date.

This Appendix to the PET assessment report presents the results of the systematic review of PET as a diagnostic test for Alzheimer's disease. Alzheimer's disease and other neurologic and psychiatric conditions are significant presences in the PET literature, and predate studies of the use of PET for diagnosis of diseases in other organ systems. Maisey and Jeffrey (1991) attribute this emphasis to the high level of metabolic activity of the brain and to the design of early PET scanners to accommodate only the head.

PET allows the qualitative and quantitative evaluation of cerebral physiology, and the exploration of the biochemical bases for clinical diseases. Fluorodeoxyglucose (FDG) PET brain studies have been used for the many research and clinical purposes related to the central nervous system, including (Hoffman, et al., 1993):

- definition of the magnitude and distribution of normal local cerebral glucose metabolism, and the effects of age and sex on metabolism;
- localization of seizure onset in patients who have partial complex seizures and who are being considered for temporal lobectomy (FDA approved use of FDG);
- assessment of brain tumors, including the degree of malignancy at the time of diagnosis, persistent postoperative tumor, differentiation of high- from low-grade tumors and radiation necrosis from persistent tumor;
- evaluation of schizophrenia, affective disorders, obsessive-compulsive disorder;
- study of cerebral metabolism in cerebrovascular disease;
- definition of regions of changed glucose metabolism in various forms of dementia, including Alzheimer's disease, Pick's disease, and Huntington's disease.

I. BACKGROUND

A. Description

Alzheimer's disease (AD) is a progressive neurodegenerative disorder, and is the most common form of dementia. Dementia is defined by the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) as "the decline in memory and other cognitive functions in comparison with the patient's previous level of function as determined by a history of decline in performance and by abnormalities noted from clinical examination and neuropsychological tests" (Morris, 1994). Dementia is a diagnosis based on behavior and cannot be determined by imaging studies or laboratory tests, although specific causes of dementia may be identified by these means. More than 55 illnesses, some nonprogressive, can cause dementia. AD, alone or in combination with other illnesses, accounts for approximately 70% of cases of dementia in industrialized countries (Geldmacher and Whitehouse, 1996).

In AD, intellectual ability, abstraction, judgment, memory, language, and finally motor functions deteriorate (Mazziotta, et al., 1992). Cerebral (brain) tissue damage is widespread and complex, with progressive loss of synaptic (intercellular) connections and cell death. Genetic linkages with a number of chromosomes, including 1, 14, 21 (early onset disease), or 19 (late onset), have been identified for familial forms of AD (FAD). However, the etiology of most forms of non-familial ("sporadic") AD remains unknown (Schorderet, 1995).

B. Epidemiology

AD is the most common cause of progressive intellectual failure in middle or late life, and is the fourth leading cause of death in the developed world (Duff and Hardy, 1994). The etiology of AD remains undefined; risk factors that have been tested in analytic epidemiology studies include family history, head trauma, aluminum exposure, and viruses. Findings on all of these potentially associated factors have been equivocal (Larson, et al., 1992).

The prevalence of AD rises steadily from late middle age in all populations that have been studied: studies using formal clinical criteria for AD (see sections on description, above, and diagnosis, below) from both Europe and the United States found rates per 100 population of 3.1 to 15.3 in individuals over 65 years of age, 4.1 to 6.1 in those over 75 years of age, and 7.1 to 47.2 in those over 85 (Rockwood and Stadnyk, 1994). AD patients have a median survival of eight to ten years after onset (range, 1 - 20 years) (Larson, et al., 1992). Over 2 million people in the United States are incapacitated by AD to the degree that they require assistance with daily living. Estimates of the cost of care for those in the United States with AD have ranged from $44 billion (Mazziotta, et al., 1992) to $100 billion (Post, 1994) per year.

As the proportion of elderly individuals in the United States increases, AD becomes an increasingly important public health concern. At present, there are more than 8 million veterans ages 65 and older (37% of the total veteran population). Improving the diagnosis and treatment of AD is a major goal of the Veterans Administration health care system (Respess, 1995).

C. Diagnosis

Much of the information in this section was taken from the chapter on dementia (McCormick and Larson, 1991) in the book *Diagnostic Strategies for Common Medical Problems*, published by the American College of Physicians. This book is intended to give practicing clinicians tools for the quantitative interpretation of clinical and diagnostic test information (i.e., tools for the evidence-based application of diagnostic tests).

In the diagnostic strategy for AD, the presence of dementia is first determined, and then its cause is established. Screening for dementia involves tests such as the Mini-Mental State Exam, which tests a broad range of cognitive functions. Once screening has documented the presence of dementia, causes other than AD are excluded. In some patients meeting the criteria for dementia, cognitive impairment is due to medication side effects, depression, other central nervous system diseases or metabolic abnormalities. Some of these may be treated, resulting in improved or stabilized cognitive function.

A definitive diagnosis of AD is based on a typical clinical picture and histopathologic findings in samples of brain tissue. The histopathologic hallmarks of AD are neuritic or senile plaques (large extracellular protein deposits) and neurofibrillary tangles (bundles of abnormal protein filaments inside nerve cells). Nerve cell damage and death is most severe in the region of the hippocampus (a deep-lying structure in the temporal lobe of the cerebral hemispheres that is involved in memory functions).

In the absence of histological confirmation of AD, patients are referred to as having a diagnosis of dementia of the Alzheimer type (DAT), rather than as having AD. Subsequent sections of this review will use the classifications “AD” and “DAT” literally: “AD” will refer to cases in which the disease has been definitively diagnosed by histopathology, while “DAT” will refer to cases to which clinical criteria only have been applied.

The prevalence of any cause of dementia varies among populations. Estimating pretest probability of disease requires consideration of multiple factors: the patient’s age and race, whether (and in what type of hospital and on which ward) the patient is hospitalized. Reviews of dementia prevalence have included the following data for the veteran inpatient population (McCormick and Larson, 1991):

Alzheimer’s disease, 49 - 70% of cases of dementia
multi infarct dementia, 7 - 22%
infection, 1 - 3%
metabolic condition, 2%
neoplasm, 1 - 5%
normal pressure hydrocephalus, 2 - 5%
subdural hematoma, 3%
depression, 3%

Huntington’s disease, 1%
Parkinson’s disease, 4%
Alcoholism, 3 - 8%
Other, 2%
(progressive supranuclear palsy, frontotemporal dementia, Pick’s disease, cortical basal degeneration)

Table 1 NINCDS-ADRDA criteria for clinical diagnosis of Alzheimer's disease

Diagnosis	Criteria	Features consistent with diagnosis	Features inconsistent with diagnosis
PROBABLE Alzheimer's disease	• dementia established • deficits in two or more areas of cognition • deterioration is progressive • no disturbance of consciousness • onset between ages 40 and 90, most often after age 65 • no other systemic disorder	• progressive deterioration of individual cognitive functions • impaired activities of daily living and altered patterns of behavior • family history of similar disorders • laboratory results of: - normal lumbar puncture - normal pattern or nonspecific changes in EEG - evidence of progressive cerebral atrophy on CT After exclusion of causes of dementia other than Alzheimer's disease: • plateaus in the course of disease • associated psychiatric symptoms, physical outbursts, sexual disorders, and/or weight loss • other neurologic abnormalities (including motor), especially with more advanced disease • seizures in advanced disease • CT normal for age	• sudden onset • focal neurologic findings • seizures or gait disturbances at the onset or very early in the course of the illness
POSSIBLE Alzheimer's disease	• dementia syndrome in the absence of other neurologic, psychiatric, or systemic disorders sufficient to cause dementia • diagnosis may be made in the presence of second systemic or brain disorder not thought to be the cause of the dementia • should be used in research studies when a single, gradually progressive severe cognitive deficit is identified in the absence of other identifiable cause		
DEFINITE Alzheimer's disease	• clinical criteria for probable Alzheimer's disease and • histopathologic evidence obtained from a biopsy or autopsy		
Classification of Alzheimer's disease for research purposes	Should specify features that may differentiate subtypes of the disorder: • familial occurrence • onset before age 65 • presence of trisomy-21 • coexistence of other relevant conditions such as Parkinson's disease		

Adapted from McKhann, et al., 1989

Clues that conditions other than AD may be the primary cause of dementia have been codified in clinical diagnostic criteria, including those of the National Institute of Neurologic and Communication Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS/ADRDA) (McKann, et al, 1984; Table 1). Diagnostic and Statistical Manual (DSM-IIIR) criteria are also used. Table 2, adopted from McCormick and Larson (1991) and Kukull, et al. (1990), lists the operating characteristics of screening tests and clinical criteria in the diagnosis of dementia and AD. The data in the table come from evaluations of the clinical criteria against the gold standard of histopathologic diagnosis.

Table 2 Tests for dementia and AD

Test		*Sensitivity (%)*	*Specificity (%)*	*Likelihood Ratio*	
				positive	*negative*
Screening for dementia	*MMSE*	87	82	4.8	0.16
Diagnosing AD in demented patients	*NINCDS/ADRDA criteria*	92	65	2.6	0.12
	DSM-IIIR	76	80	3.8	0.30

The information in Table 2 should be considered when interpreting the published evaluations of PET's diagnostic accuracy in AD. Kukull, et al. (1991), point out that investigators wishing to ensure that patients classified as AD are more likely to be AD should choose DSM criteria, while investigators wishing to include the greatest number of AD cases, seldom assigning a false-negative diagnosis, should choose NINCDS/ADRDA criteria. Gearing, et al. (1995) note that diagnostic accuracy has improved over time with the increasing use of formal clinical criteria and that in none of the first 106 autopsies in demented patients enrolled in the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) study was a potentially treatable disorder mistakenly diagnosed as AD using clinical criteria.

D. Treatment

No cure for AD is currently available; two drugs (tacrine and donepezil) that appear to modify the course of AD in some patients have been approved by the FDA. Pharmacotherapy is also used to treat some of the neuropsychiatric and behavioral disturbances associated with AD, and pharmacologic agents intended to affect AD cognitive dysfunction directly are under investigation. The conceptual framework for these treatments is that AD is a progressive degenerative dementia caused by the loss of neurons, synapses, and associated metabolic dysfunction. Therefore, treatment efforts are focused on the replacement or enhancement of the function of existing neurons (Whitehouse and Geldmacher, 1994).

Several neurotransmitter systems affected in AD may contribute to the cognitive dysfunction and provide the basis for neurotransmitter replacement therapy. Dysfunction of the cholinergic system in AD and evidence that this system is involved in human cognition have led to clinical trials of cholinergic agents. Tacrine and donepezil, cholinesterase inhibitors, are currently the only FDA-approved drugs for the treatment of AD. Tacrine has a modestly positive effect on cognitive and behavioral function in 30% to 50% of mild to moderately impaired AD patients (Davis, et al., 1992); the clinical significance of these effects has been questioned, and tacrine does not affect the course of the disease (Growdon 1992; Crismon, 1994). Tacrine has significant side effects, including reversible liver damage and cholinergic adverse effects (nausea, vomiting, diarrhea, abdominal pain, dyspepsia) (Growdon, 1992; Wagstaff and McTavish, 1994; Whitehouse and Geldmacher, 1994). Donepezil does not appear to be associated with hepatotoxicity (Rogers, et al., 1996).

E. Rationale for PET in AD

The primary role of diagnostic testing has been differential diagnosis of AD from reversible or treatable diseases. These include dementia due to medication intoxication, infection, metabolic or nutritional disorders; benign brain tumors; normal pressure hydrocephalus; or multiple infarct dementia (MID) due to a series of small strokes (Kuhl, 1991).

As discussed above, the clinical diagnosis of DAT does not correspond to AD in 100% of cases; diagnosis early in the course of the disease can be particularly problematic (Hoffman, 1993). Initial studies into the use of PET in patients who met clinical criteria for DAT were based on the desire to improve diagnostic certainty and to provide information on the pathophysiologic basis of the disease. With the availability of tacrine and donepezil and the ongoing research into other drug therapies for AD, a renewed impetus for an accurate clinical diagnosis, including a diagnosis for the very early stages of dementia, has been noted (Morris, 1994).

Jobst, et al. (1994), discussing CT as a diagnostic test for AD, summarize the reasons to work toward increasing the accuracy of antemortem AD diagnosis. While treatment of AD is still at a rudimentary stage, accurate diagnosis is a prerequisite to the selection of defined cases for evaluation of therapies. There is also a need for precise epidemiologic and demographic knowledge about Alzheimer’s disease and other dementias and for the best possible antemortem diagnosis so that patients and their families can be provided with clear information that enables them to organize their lives.

F. Special methodologic considerations in evaluating a diagnostic test for Alzheimer’s disease

Accurate estimation of the characteristics of a diagnostic test depends on the test’s comparison with a (hypothetically) 100% accurate “gold standard” test. When PET is used to diagnose or stage cancer, as in the other systematic reviews conducted for this assessment, its results are compared to those obtained by biopsy. Biopsy, while imperfect, generally offers a quite accurate estimate of the presence or absence of cancer. Clinical criteria have been shown to be less than optimally accurate when compared to biopsy or autopsy diagnosis of Alzheimer’s disease, and the potential biases discussed below can be assumed to be operating in the studies of PET as a diagnostic test for DAT.

If the reference or gold standard test is significantly inaccurate, the estimates of the new test’s characteristics will be biased. Since the results of the new test and the reference test are likely to be positively correlated (i.e., statistically dependent), this bias will result in higher estimates of sensitivity and specificity for the new test than would be the case if the test were compared with a true measure of disease status. In the rarer cases where the reference test and the new test have statistically independent results, results with the new test will be biased toward zero for both sensitivity and specificity (Phelps and Hutson, 1995).

Methods for correcting bias in estimates of a new test's accuracy compared to an inaccurate "gold standard" test have been developed. Some of the methods have focused on the assumption that the reference test and the new test have statistically independent results, and permit an algebraic correction under specific circumstances (e.g. the reference test has known sensitivity and specificity) (Begg, 1987). However, the assumption of statistical independence of the new and reference tests is usually quite implausible for most applications, since covariates of both test results (such as stage or severity of disease) will affect both tests simultaneously (Begg, 1987).

Other, recently developed methods are applicable when the two tests are either statistically dependent or independent (Phelps and Hutson, 1995). These methods require the research studies measuring the new test's accuracy to estimate a *probability* that each subject is abnormal with the gold standard process, rather than a binary (normal vs abnormal) measure. Phelps and Hutson provide an example of the application of these methods to the use of MRI for the diagnosis of multiple sclerosis (MS).

In the absence of studies specifically designed for use of the correction methods outlined above, Begg (1989) recommends that at least a subset of patients in each diagnostic accuracy study for a new test have been definitively diagnosed. The estimates obtained from studies where most patients have not been definitively diagnosed must be interpreted with reference to the imperfect standard.

Ideally, evaluation of the accuracy of a diagnostic test for AD should rely on data obtained from cohort studies in which the test is applied at intervals before death in patients with DAT, other forms of dementia, and in controls, and all subjects are followed to death and autopsy. Such studies have been performed for CT and SPECT (see below), but the MDRC Technology Assessment Program was unable to locate any published studies that had used similar methods to evaluate PET.

A cooperative group of European PET centers *is* currently conducting such a study, which incorporates a standardized neuropsychological test battery, standardized PET data analysis, and follow-up to autopsy with standardized neuropathologic criteria (Dr. K. Herholz, Max Planck Institut, Germany; personal communication, 1996). The study focuses on patients with NINCDS/ADRDA "possible" AD (i.e. the group of patients in whom there is the greatest uncertainty regarding diagnosis and for whom a more accurate test would most contribute to posttest certainty) and patients with other causes of dementia. Copies of the study protocol are available from the MDRC Technology Assessment Program.

G. Alternative neuroimaging technologies and other tests relevant to diagnosing AD

1. Neuroimaging technologies

Neuroimaging (i.e., CT) for cases of suspected AD is generally considered only after a systematic evaluation of a patient's mental status and history, and of information from reliable informants. Diagnostic criteria help to rule out other causes of dementia. Based on the prevalence figures for AD in the hospitalized veteran population cited above (approximately 50% to 70%), clinical criteria (likelihood ratio of 1.3 to 2.8 in the presence of a positive test result, from Table 2) give posttest probabilities of disease from 75% to 90%. The American College of Physicians (McCormick and Larson, 1991) recommends that CT be reserved for patients in whom there is a clinical suspicion of a focal or destructive central

nervous system lesion (i.e., patients whose neurological exam or clinical history is more suggestive of a focal central nervous system lesion than of AD or other causes of dementia such as Parkinson's disease).
Jobst, et al. (1994) confirm that the role of neuroimaging in AD has been to identify and exclude other intracerebral pathologies. The results of current work with both PET and other neuroimaging technologies suggest that neuroimaging may eventually play a more direct diagnostic role. Alternate neuroimaging technologies, both structural (CT and MRI) and functional (SPECT), have generated equivalent levels of research activity to that seen in support of PET as a diagnostic test for AD. These technologies are generally more widely available than PET, and if diagnostic accuracy were comparable, would be likely to be more widely used.

Neuroimaging alternatives that have been directly compared to PET in cross sectional studies using clinical criteria for DAT as the diagnostic standard include MRI, CT, and SPECT (Table 5). CT, SPECT, and the combination of CT and SPECT have been studied in cohorts of patients followed to autopsy (Jobst, et al., 1992 and 1994; Table 6); these studies use definitive diagnosis by histopathology as the gold standard, and provide estimates of sensitivity and specificity unbiased by the inaccuracies associated with clinical criteria. The studies in Table 6 also document diagnostic thinking efficacy: Jobst, et al., provide information allowing the calculation posttest probability of disease from likelihood ratios and age-specific pretest probabilities; and Van Gool, et al., document the incremental contribution of SPECT after careful clinical and laboratory examinations in mildly demented elderly patients.

The studies in Table 6 provide models that would benefit all evaluations of tests for AD. Their methodological strengths include:

- cohort design incorporating an exceptionally high (96%) rate of consent to autopsy [using methods documented by King, et al. (1993)] in AD cases, cases with other dementing conditions, and controls (Jobst, et al., 1992 and 1994);
- results are framed in clinically useful terms (providing age-related pretest probability of disease and likelihood ratios from which to calculate posttest probability of disease) (Jobst, et al., 1992 and 1994);
- the interrater reliability of the tests has been calculated (Jobst, et al., 1992 and 1994);
- control groups include both those without dementia and those with other dementias (Jobst, et al., 1992 and 1994; Van Gool, et al., 1995);
- since discrepancies among all the criteria sets used for histopathologic diagnosis of AD are well recognized (Jarvik, et al., 1995), all diagnoses are made with rigorous application of the same criteria set (Jobst, et al., 1992 and 1994);
- the incremental diagnostic certainty supplied by an additional test is defined (Van Gool, et al., 1995).

2. Other tests

Other diagnostic tests are also under development for AD. These include: presence of the ε4 allele for apolipoprotein E (Nalbantoglu, et al., 1994; Reiman, et al., 1996; National Institute on Aging/Alzheimer's disease, 1996); tau (microtubule associated) protein in cerebrospinal fluid (Arai, et al., 1995); decreased ß-amyloid peptide$_{42}$ in cerebrospinal fluid (Motter, et al., 1995), and hypersensitivity of pupil responses to tropicamide (Scinto, et al., 1994). While the these tests and other tests are still under investigation, their potential use could have important implications for the role of more expensive and less widely available technologies such as PET.

H. Ethical considerations in testing for AD

Technical efficacy studies (Section VII) have used PET to identify metabolic changes in the brains of patients with early DAT; pre-symptomatic individuals at risk for familial AD have also participated in PET studies (Small, et al., 1995). Other means of identifying individuals at risk for AD, such as apo ε typing, are commercially available (National Institute on Aging/Alzheimer's Association Working Group, 1996).

The National Institute on Aging/Alzheimer's Association Working Group (1996) noted that genetic risk factor assessment applied to diseases, such as AD, that involve the interaction of several genes and environmental factors is complicated by uncertainties in predicting and diagnosing multifactorial disorders, and by the potentially far-reaching social, ethical, and medicolegal implications of disclosure of genotype results. Many of these uncertainties would complicate any test for early or preclinical AD.

Post (1994) summarized ethical issues in AD, some of which relate to early diagnosis:

- pre-emptive suicide may be considered by patients who have received a diagnosis of AD early in the course of the disease;
- early diagnosis, in the absence of interventions to modify risk or treat the disease, may be associated harms that outweigh its benefits;
- patients have a legal and ethical right to decide, while still competent, to use or reject technologies should they become incompetent
- independent living, driving, insurance, and jobs may be threatened by a diagnosis of AD.

II. RESULTS

Fifty-five articles were selected from MEDLINE and other database searches and from the bibliographies of initially retrieved articles as meeting the screening criteria. After review, 23 (42%) were found to meet inclusion criteria: 15 met the definition of technical efficacy (Fryback and Thornbury, 1991; *Appendix 2: Assessing Diagnostic Technologies*); 6 met (with the exception of the gold standard) the evidence-based criteria for studies of a single diagnostic test, and an additional 2 studies met the evidence-based criteria while comparing PET with other neuroimaging tests.

Technical efficacy studies are listed in Section VII, below; data abstraction tables for these studies are on file with the MDRC Technology Assessment Program. For this review, the definition of technical efficacy was expanded to include those not designed to assess diagnostic accuracy or not meeting the evidence-based criteria for diagnostic accuracy. These studies do provide information necessary to subsequent diagnostic efficacy studies (Table 4). The MDRC Technology

Assessment Program was unable to locate any published PET studies at higher levels of the Fryback and Thornbury diagnostic efficacy hierarchy.

The technical efficacy studies listed in Section VII compared patients with DAT to non-demented controls and tested the differences between groups with inferential statistics. Rapoport (1991) summarized some of the conclusions that had been drawn at that time regarding brain metabolism in DAT from the studies that met technical efficacy criteria for this review and from other studies that have appeared in the literature:

- reductions in resting state regional brain metabolism are roughly proportional to dementia severity;
- metabolic reductions are greater in association areas than in primary sensory and motor neocortical areas, and correlate with the distribution of neuropathology and cell loss postmortem;
- brain metabolic patterns in DAT patients are heterogeneous, belonging to at least four distinct metabolic groups that correspond to different patterns of cognitive and behavioral abnormalities;
- abnormal left/right asymmetries in mild DAT can retain their initial direction for extended periods, and may precede and predict the cognitive deficits that later appear;
- parietal association/frontal association metabolic ratios also retain their direction over time;
- although metabolically spared compared to the association cortices, the primary sensory cortices, basal ganglia, thalamus, and cerebellar hemispheres show metabolic declines over time with high resolution scanners.

Table 4 presents the diagnostic accuracy efficacy studies located by the MDRC Technology Assessment Program. Since these studies used cross sectional with controls design, posttest probability of disease can be calculated and the studies can also be classified at the diagnostic thinking efficacy level. Additional support for the ability of PET to accurately predict clinical classification of DAT is provided by Hoffman, et al. (1996), who found that FDG PET studies had high inter- and intra-rater reliability in distinguishing probable and possible AD from other potential causes of dementia and memory disturbance.

III. SUMMARY

Table 3 summarizes published findings on the diagnostic accuracy and diagnostic thinking efficacy of PET and its neuroimaging alternatives. The PET studies did not met evidence-based medicine criteria (since histopathologic diagnosis was not the gold standard) but otherwise fulfilled most requirements for good methodologic quality. Unique features of each test are noted, as are the comparison groups used in each study. PET results were not quantitatively pooled, as each published study used a different method for PET data analysis and accuracy results fell within relatively narrow ranges. Histopathology is recognized as the gold standard for diagnosing Alzheimer's disease; studies that evaluate PET against clinical criteria may overestimate accuracy.

Table 3 Diagnostic accuracy and diagnostic thinking efficacy of PET and its neuroimaging alternatives

Neuroimaging Test	Diagnostic Standard Used in Evaluation Studies		Characteristics
	Histopathology	*Clinical criteria*	
CT	x		Se = 94%; Sp = 93.5% (AD-specific orientation; AD vs normal controls and other dementias) *Jobst, et al., 1994*
SPECT	x		Se = 96%; Sp = 89% (AD vs normal controls and other dementias) *Jobst, et al., 1994*
		x	Sp = 89% • all probable AD, Se = 43% • probable AD < 80 years, Se = 56% • probable AD > 80 years, Se = 29% • SPECT contributed to 8% of final diagnoses *Van Gool, et al., 1995*
CT + SPECT	x		Se = 90%; Sp = 97% (AD vs normal controls and other dementias) *Jobst, et al., 1994*
PET		x	Se = 94.6; Sp = 97% ("robust ratio"; DAT vs normal controls) *Herholz, et al., 1993* Post test probability of disease, positive test = 90%; posttest probability, negative test = 10% in patients with pretest probability of disease = 50% (neural net; DAT vs normal controls) *Kippenhan, et al., 1994* Se = 94%; Sp = 79% (4 image patterns typical of DAT; DAT vs normal controls) *Salmon, et al., 1994* Se = 94%; Sp = 53% (4 image patterns typical of DAT; DAT vs non-DAT dementia controls) *Salmon, et al., 1994* Se = 94%; Sp = 99% (stereotactic surface projections; DAT vs non-DAT controls) *Burdette, et al., 1996*
PET vs CT		x	PET: Se = 97%; Sp = 84% (qualitative) CT: Se = 86%; Sp = 28% (cortical atrophy) (DAT vs normal controls) *Fazekas, et al., 1989*
PET vs MRI		x	PET: Se = 97%; Sp = 84% (qualitative) MRI: Se = 92%; Sp = 60% (ventricular atrophy) (DAT vs normal controls) *Fazekas, et al., 1989*
PET vs SPECT		x	PET: Se = 80%; Sp = 100% (typical functional pattern) SPECT: Se = 80%; Sp = 65% (typical functional pattern) (DAT vs normal controls and vascular dementia) *Mielke, et al., 1994*

Se = sensitivity; Sp = specificity

IV. DISCUSSION

The face value of PET's diagnostic accuracy in AD appears to be very good, and fairly equivalent across a variety of data analysis methods and scanning protocols (Table 3; Herholz, et al., 1993). However, PET has been evaluated against clinical criteria (an imperfect diagnostic standard) only. Since the factors that affect clinical criteria accuracy (e.g. severity of disease) also are likely to affect PET results, the published sensitivity and specificity figures may be overestimates. The discussion sections of these papers note that further studies comparing PET to definitive diagnosis by histopathology are necessary to confirm results; a large cooperative study in Europe using histopathology as the diagnostic standard is currently under way.

An additional source of bias may be attributable to the choice of clinical criteria in PET diagnostic accuracy studies. All used NINCDS/ADRDA criteria, which are associated with a higher rate of false positives and a lower likelihood ratio for a positive test (2.6) than are the DSM-III criteria (likelihood ratio = 3.8). If DMS-III criteria are applied in patients with a pretest probability of AD of 60% (the midpoint of the hospitalized veteran prevalence range), the posttest probability of AD is approximately 85%. NINCDS/ADRDA criteria yield an approximately 80% posttest probability of AD.

Finally, most of the cases in PET diagnostic accuracy studies had possible or probable AD according to NINCDS/ADRDA criteria. Few studies applied PET prospectively to large numbers of patients with other diagnoses (e.g., vascular dementia), which would be necessary to fully define the positive predictive value of PET as a diagnostic test.

The clinical importance of differences in clinical criteria accuracy and the additional accuracy attributable to PET [posttest probability of disease in the hospitalized veteran population with face value sensitivity and specificity from Herholz, et al. (1993) of > 99%] rests on changes in management or treatment decisions that follow test results. Recent studies indicate that as clinicians have gained experience in the application of clinical criteria their accuracy has increased, and that treatable causes of dementia are rarely missed. Since treatment options for AD itself are currently limited and use of clinical criteria appears to miss very few treatable causes of dementia, increased diagnostic accuracy may be needed primarily in research settings (epidemiologic studies and evaluations of potential therapies). The value of improved diagnostic information to patients and their families should not be dismissed; however, this value remains unquantified.

The accuracy and potential research utility of PET in AD should be viewed in the context of the accessibility and accuracy of other imaging technologies (Tables 5 and 6), and that of other tests that are currently available or under development. In studies directly comparing PET with standard CT, MRI, or SPECT using clinical criteria as the diagnostic standard, PET has superior characteristics (Table 5). On the other hand, AD-specific and relatively simple CT and SPECT methods have been tested in rigorously designed cohort studies in which demented patients (AD and other causes) and non-demented controls have been followed to death and definitive diagnosis; these studies indicated that CT and SPECT may have sensitivity and specificity close to that of PET (Table 6).

V. CONCLUSIONS: Clinical use of PET in Alzheimer's disease

As of September, 1996, the accuracy of FDG PET in diagnosing Alzheimer's disease had been demonstrated in 5 published studies that used a variety of methods to analyze PET data and to arrive at decisions about the presence or absence of disease. These studies compared PET to clinical criteria for dementia of the Alzheimer's type. While the clinical criteria are known to be somewhat inaccurate in diagnosing AD, compared to the gold standard of histopathologic diagnosis, careful application of the criteria does appear to identify most cases of treatable dementia.

The factors noted in the discussion section and the paragraph above argue that routine clinical application of PET as a diagnostic tool for AD should await the results of the ongoing European multicenter study that will evaluate PET's accuracy against the diagnostic standard of histopathology, as well as development of more effective treatments and risk modification interventions for AD. The multicenter study's results will allow more explicit comparisons among PET and more widely available tests that may have comparable accuracy. In the absence of effective treatments for Alzheimer's disease, an accurate diagnostic test may be needed primarily in the efficacy of treatment research setting.

Finally, the role of diagnostic tests in treatment efficacy research protocols may need refinement as subsets of Alzheimer's disease are defined by phenotypic or genetic markers. Nalbantoglu, et al. (1994) calculated the population attributable risk for the ε4 allele for apolipoprotein E at about 50%. This level of attributable risk suggests that late-onset AD consists of at least two disease entities with separate underlying causes or aggravating factors. Each of the disease entities may require different treatment strategies, and selection of patients for treatment may be based on molecular or genetic tests, rather than on anatomic or functional imaging studies. Small, et al. (1995) found that the ε4 allele is associated with reduced cerebral parietal metabolism and increased asymmetry in non-demented relatives at risk for probable AD. Measurement of glucose metabolism using PET could be a means of monitoring experimental treatment responses during early phases of AD.

Table 4 Summary of the literature: Diagnostic accuracy efficacy of FDG PET imaging in Alzheimer's disease

Notes: *Studies are listed in order of date of publication. Where substantial duplication in purpose of study, patients studied, and results in multiple studies from the same institution could be inferred, only the most recent, largest, most rigorously designed, or most comprehensive was included in the table. Studies reviewed but not included are listed under "References".*

All studies in this table used a cross sectional design with controls, and provided Level III evidence (see Table 2 for an explanation of levels of evidence).

In several studies in this table, images were interpreted without blinding of image readers to clinical diagnosis, or blinding was not noted. In these studies, however, PET data were analyzed quantitatively (often by automated processes) rather than visually (qualitatively), minimizing observer bias.

Study	*Subjects/Methods*	*Results/Comments*
Azari, et al., 1993 *NIA/NIH, NIMH/NIH, Washington State Mental Illness Research & Training Institute*	**Purpose** to investigate whether a multiple regression/discriminant analysis procedure would distinguish mildly/moderately demented patients with probable AD from controls **Cases** 19 mildly/moderately demented patients with NINCDS-ADRDA probable AD **Subject at risk** 1 subject at risk for familial AD with only delayed memory at time of study **Controls** 22 healthy age- and sex-matched controls **Methods** • PET scans parallel to IOM obtained • absolute and normalized GMR obtained for 65 ROIs and whole brain • analysis involved: - selection of 2 sets of regions as dependent variables (frontal/parietal association areas and 4 smaller ROIs) - stepwise multiple regression to identify best predictors - application of regression weights to control and AD data - application of discriminant analyses to determine the weighting of regression residuals to maximize differences between cases and controls - cross validation of discriminant functions using jackknife procedure • each subject classified as control or AD using a discriminant function • discriminant functions applied to subject at risk	**Estimated posttest probabilities of group membership:** • using statistical functions based on frontal/parietal region, 95% of AD cases and controls correctly classified • average probability of correct classification: - cases, 0.97 (0.78 - 1.00) - controls, 0.93 (0.53 - 1.00) **Cross validation of discriminant functions, frontal/parietal region:** • 89% of AD patients and 86% of controls correctly classified • average probability of correct classification: - cases , 0.90 (0.55 to 1.00) - controls, 095 (0.52 - 1.00) **Cross validation of discriminant functions, 4 smaller ROIs:** • 88% of AD patients and 81% of controls correctly classified • average probability of correct classification: - cases,1.00 - controls, 0.98 (0.83 - 1.00) **Application of discriminant functions to subject at risk:** • frontal/parietal: probability of classification as AD, 0.74; as control, 0.26 at first PET study; probabilities 0.84 and 0.16, respectively, at second PET study (1 year after first) • 4 smaller areas: first and second PET studies showed had probability of classifying subject as AD of 1.00 **Success of group separation** 2 discriminant functions separated groups (5 subjects misclassified) with less overlap than single, normalized glucose metabolic index. **Conclusion** This statistical approach may be useful for early detection of AD. **Study design limitation** no definitive diagnosis in any DAT subjects

Study	*Subjects/Methods*	*Results/Comments*
Herholz, et al., 1993 *Max Planck Institut, Cologne; Hospital San Raffaele, Milan; Université de Liège*	**Purpose** to assess whether a study protocol based on a robust ratio to assess the typical metabolic pattern of AD can yield comparable results in 3 different centers **Cases** 37 patients with NINCDS-ADRDA probable AD **Controls** 34 healthy subjects **Methods** • multiple PET slices parallel to CM line from cerebellum to ≥ 27 mm above basal ganglia • GMR calculated according to machine and software properties of each center • experienced physician performed examined images visually, with access to clinical information (analogous to clinical practice) • ROIs defined and GMRs calculated for regions most typically affected by AD (temporoparietal and frontal) • composite metabolic ratio representing typical pattern in AD calculated	**Visual analysis** • high frequency of typical pattern (bilateral temporoparietal and optional frontal hypometabolism) **Metabolic ratio analysis** • AD metabolic ratios significantly lower than controls • differences among ratio means at centers not significant • ratio increased significantly with age, but centers did not differ after age adjustment of ratios • mean value of ratio close to 1.0 over entire age range in controls **Diagnostic accuracy** • composite metabolic ratio yielded better separation of cases from controls than did ratios of single regions by ROC curve analysis • at cutpoint of 0.921, Se = 94.6% and Sp = 97% (95.8% of subjects correctly classified) **Factors affecting variation among centers** • rate constants used (10% change in constant produced < 1% change in composite ratio) • region size (10% increase in size increased ratio by 1.2%) **Conclusion** A common investigation protocol may yield comparable PET data from different centers in spite of differences between scanners and imaging equipment. **Study design limitations** no definitive diagnosis in any DAT subjects
Kippenhan, et al., 1994 *University of Miami , NIA/NIH*	**Purposes** • to generate recommendations for optimal data representation and analysis in diagnosis of AD • to compare the ability of 2 PET cameras (PETT V and Scanditronix, a higher resolution scanner) to diagnose AD using optimal discriminators and a neural network • to define the most generally applicable metabolic discriminators of AD **Cases** • PETT V: 41 patients with NINCDS-ADRDA probable AD • Scanditronix: 33 patients with NINCDS-ADRDA probable AD **Controls** • PETT V: 50 age-matched normal individuals • Scanditronix: 74 age-matched normal individuals **Methods** • small structures from Scanditronix database combined to obtain regional representations equivalent to those of PETT V data at lobular and lobar levels • classification performance from each database evaluated for lobular and lobar representations for various methods of classification (by ROC analysis) and data processing • classifiers evaluated by cross-validation studies on training and testing sets • neural network training by back-propagation techniques • classification results for neural net compared to results using discriminant analysis • different methods to preprocess data compared	**Classification according to global metabolism** areas under ROC curve: Scanditronix .90, PETT V .60 **Results of optimization experiments** • lobular representation and occipital normalized data resulted in best performance for PETT V • lobular data processed with either simple scaling or normalization **Comparison of neural net and discriminant analysis** • performance approximately equal for Scanditronix lobular data • performance of neural net somewhat higher for PETT V lobular data **Can neural nets identify groups in one database after being trained with sets including subjects from other database?** better performance can be expected by training with lower resolution data ("noisier") and testing on higher resolution data than the reverse (for normalized data) **Most important and generalizable discriminating profiles learned by neural nets during lobule-level training with both databases** • generally low metabolic values in parietal and temporal areas • higher values in motor-sensory and occipital regions • asymmetry **Posttest probabilities of disease** • classification at point of maximum information on ROC curve for normalized Scanditronix lobular data resulted in posttest probability of 90% (rule in disease) for an abnormal test and 10% (rule out disease) for a negative test • corresponding posttest probabilities for PETT V lobular data were 87% and 24% **Authors' comment** it should be possible to share metabolic data from different scanners and institutions to develop an extensive knowledge base of metabolic patterns **Study design limitation** no definitive diagnosis in any AD cases

Study	Subjects/Methods	Results/Comments
Salmon, et al., 1994 *University of Liège, Belgium*	**Purpose** to evaluate the role of visual analysis of PET metabolic patterns in patients referred for differential diagnosis of degenerative dementias **Cases** 65 patients with NINCDS-ADRDA probable AD (5 with a diagnosis of definite AD after histologic examination) **Controls** 64 patients whose final diagnosis was: degenerative dementia atypical for AD (possible AD, 19); Parkinson's disease (13); progressive supranuclear palsy (1); vascular dementia (8); mixed dementia (9); Creutzfeldt-Jacob disease (3); metachromatic leukodystrophy (1); dementia from anoxia (1); primary progressive aphasia (2); normal pressure hydrocephalus (3); depression (4) (7/64 confirmed by histologic diagnosis) **Methods** • PET scans acquired in resting state on plane parallel to IOM line • ROIs and visual analyses performed on 7 planes selected with brain atlas • 9 categories of image patterns for hypometabolism: - bilateral temporo-parietal (with/without frontal) - unilateral temporo-parietal (with/without frontal) - frontal regions bilaterally affected more than temporo-parietal - frontal unilaterally affected more than temporo-parietal - isolated bilateral frontal involvement, sometimes asymmetrical - left perisylvian, sometimes extending to homolateral cortices - diffuse cortical, localized above level of basal ganglia - multiple patchy foci, cortical and subcortical - normal • reader (number not specified) blind to clinical data except for suspicion of dementia	**General findings in probable AD group** • 97% of PET scans abnormal • 2 patients with mild dementia had normal scans (1 showed AD pattern 5 years later) • metabolic pattern in AD is heterogeneous; multiple cut points corresponding to subgroups of the 9 patterns are possible **If first 4 image patterns considered positive for AD** • distinguishing AD from dementia atypical for AD: Se = 94%; *Sp = 79%; *PPV = 94%; *NPV = 79% • distinguishing AD from all controls: *Se = 94%; *Sp = 53%; *PPV = 67%; *NPV = 89% **If bilateral temporo-parietal hypometabolism only considered positive for AD** • distinguishing AD from dementia atypical for AD: Se = 66%; *Sp = 89%; *PPV = 96%; *NPV = 44% • distinguishing AD from all controls: *Se = 66%; *Sp = 54%; *PPV = 68%; *NPV = 52% **Study design limitations/comments** • final diagnosis obtained after unspecified time of follow up in many patients; data used at follow up included PET (i.e., diagnostic standard not applied without knowledge of PET result and possible incorporation bias) • only 19 controls (with possible AD)contributed data to authors' analyses • histologic confirmation of diagnosis in small number of patients • "Results" section difficult to interpret; authors indicate that analyses will be restricted to patients with probable or definite AD (cases) and dementia atypical for AD (19/64 controls) but give diagnostic accuracy figures based on all controls • study has significant value in that it tested PET in a population of patients with readily confused diseases (a common clinical situation)

Study	Subjects/Methods	Results/Comments
Burdette, et al., 1996 *University of Michigan, Ann Arbor*	**Purpose** to compare diagnostic accuracy of 3D stereotactic surface projection PET images to accuracy of transaxial images using ROC analysis in cases and controls obtained retrospectively from research database **Cases** 39 patients with NINCDS-ADRDA probable AD • 28 questionable/mild dementia • 11 moderate/severe dementia **Controls** 40 patients without dementia • 18 patients with cerebrovascular disease (5 multiple infarctions, 8 cerebrovasculitis with systemic lupus erythematosis, 5 moderate/large vascular distribution infarction) • 22 similar-aged normal individuals from database **Methods** • 2 sets of PET images (ordinary projection and 3D stereotactic surface projection) analyzed qualitatively by 2 expert and 2 novice interpreters who received brief training sessions on day of image analysis • each image set presented to interpreters in different randomized order • images scored: definite AD; probable AD; indeterminate; probably not AD; definitely not AD • interpreters blinded to patient identifiers and clinical information • interpreters' response data analyzed using ROC techniques (area under curve, SD, two-tailed p value for each reader and each type of presentation) • data from questionable/mild dementia (more difficult diagnosis) analyzed separately • Se, Sp calculated using definite and probable scores as positive for AD **Study design limitation** no definitive diagnosis in any AD cases	**Discrimination of AD using ΔZ-score, previous study** (Minoshima, et al., 1995) at Sp = 100% (zero false positives): • parietal (cutpoint ΔZ = .55) Se = 95% • temporal (cutpoint ΔZ = .45) Se = 81% • frontal (cutpoint ΔZ = .56) Se = 59% • unilaterally averaged ΔZ (cutpoint ΔZ = 0.36) Se = 97% • bilaterally averaged parietal-temporal-frontal ΔZ (cutpoint ΔZ = .52) Se = 100% • robust ratio from Herholz, et al. (table entry above, cutpoint = .89) Se = 92% • false negatives were mildly demented patients with unilateral hypometabolism • cerebrovascular disease cases: quantitative indices yielded some false positives, but distributions of metabolic abnormalities were clearly distinguishable from AD on visual inspection **ROC analysis, this study** • diagnostic accuracy improved in all readers with 3D stereotactic projections • no difference between beginner and expert readers with 3D stereotactic projections **Transaxial image display, all readers** • Se = 85% ± 0.06 (95% CI. ± 0.10) • Sp = 88% ± 0.02 (95% CI. ± 0.15) **3D stereotactic surface projection, all readers, all subjects** • Se = 94%± 0.02 (95% CI, ± 0.04) • Sp = 99% ± 0.02 (95% CI, ± 0.03) **All readers, questionable/mild dementia subjects** • transaxial projection: Se = 79% ± 0.06 (95% CI, ± 0.02) • 3D stereotactic projection: Se = 94%± 0.01 (95% CI, ± 0.02)

* = calculated by MDRC Technology Assessment Program from information provided in the article

Abbreviations:

AD, Alzheimer's disease
CI, confidence interval
CT, computed tomography
DAT, dementia of Alzheimer's type
EEG, electroencephalography
GMR, glucose metabolic rate
IOM, inferior orbitomeatal line
MANOVA, multivariate analysis of variance
MID, multi-infarct dementia
MMSE, Mini-Mental State Examination
MRI, magnetic resonance imaging
NIA, National Institute on Aging
NIH, National Institutes of Health
NIMH, National Institute of Mental Health
NINCDS-ADRDA, National Institute of Neurological and Communicative Disorders and Stroke-Alzheimer's Disease and Related Disorders Association
NPV, negative predictive value
PPV, positive predictive value
ROC, receiver operating characteristic
ROI, region of interest
Se, sensitivity
Sp, specificity

Table 5 Diagnostic accuracy efficacy of neuroimaging alternatives to FDG PET for Alzheimer's disease

Cross sectional studies with controls comparing PET to alternative diagnostic imaging tests using clinical criteria as the diagnostic standard

Notes: *Studies in this table were identified by searches of MEDLINE files for the years 1991 to 1996, using the terms "Alzheimer's disease" and "diagnosis". These studies directly compare other neuroimaging technologies with PET.*

All of the studies used cross sectional designs with controls and appeared to adequately match cases and controls for critical demographic factors.

Study	***Subjects/Methods***	***Results/Comments***
Fazekas, et al., 1989 *University of Pennsylvania*	**Purpose: PET vs CT vs MRI** to describe the type and frequency of brain abnormalities detected by CT, MRI, and PET in DAT and normal aging *information provided allowed calculation of diagnostic characteristics of each test* **AD cases** 30 DAT (DSM-III) • 24 "probable" and 6 "possible" using NINCDS-ADRDA • 14 with mild to moderate DAT according to MMSE, 16 with moderate to severe **Controls** 25 elderly individuals without evidence of dementia, and with medical findings comparable to those in DAT group **Methods** • 28 DAT and 25 controls had CT • 23 DAT and 10 controls had MRI • 30 DAT and 25 controls had FDG PET • for each imaging modality the scans from DAT and control subjects were randomly mixed and interpreted separately by a neuroradiologist (CT and MRI) and a nuclear medicine specialist (PET), without information on the age, sex, or clinical condition of patient • extent of cortical and ventricular atrophy (CT and MRI) and severity of metabolic abnormalities (PET) rated as absent, mild, moderate, or severe	**CT** • cortical atrophy: *Se = 86%; *Sp = 28%; *PPV = 57%; *NPV = 64% (mild, moderate, severe atrophy grouped as abnormal) • ventricular atrophy: *Se = 79%; *Sp = 72%; *PPV = 76%; *NPV = 75% (mild, moderate, severe atrophy grouped as abnormal) • higher grades of cortical and ventricular atrophy found in DAT than in controls and differences in mean rating between DAT and controls was significant ($p < .001$), but considerable overlap in atrophy scores **MRI** *(text and tables present different counts for DAT patients who received MRI; table number used in calculations below)* • cortical atrophy: *Se = 92%; *Sp = 10%; *PPV = 55%; *NPV = 50% (mild, moderate, severe atrophy grouped as abnormal) • ventricular atrophy: *Se = 92%; *Sp = 60%; *PPV = 73%; *NPV = 86% (mild, moderate, severe atrophy grouped as abnormal) • periventricular and/or white matter lesions: *Se = 83%; *Sp = 40%; *PPV = 63%; *NPV = 67% **PET** • *Se = 97%; *Sp = 84%; *PPV = 89%; *NPV = 95% (mild, moderate, severe hypometabolism grouped as abnormal) • abnormalities predominately focal in early DAT, diffuse hypometabolism in more advanced DAT • using visual criteria, the majority of focal metabolic abnormalities could not be explained on the basis of cortical atrophy alone; metabolic dysfunction may precede anatomic changes **Study design limitations** • protocol did not provide for image interpretation by ≥ 2 observers • no definitive diagnosis in any DAT subjects

Study	Subjects/Methods	Results/Comments
Mielke, et al., 1994 *Max Planck Institut, Germany*	**Purpose: PET vs SPECT** to define the relative ability of HMPAO SPECT and FDG PET to distinguish AD, vascular dementia (VD), and controls **AD cases** 20 patients with NINCDS-ADRDA probable AD **VD controls** 12 patients selected by modified Hachinski scores and NINDS-AIREN criteria **Normal controls** 13 normal subjects with no clinical evidence of cognitive deficits or neurological disease who were part of larger sample with subjective memory impairment **Methods** • patients received SPECT and PET on same day • SPECT and PET images coregistered and standardized ROIs generated • relative regional HMPAO uptake used to assess regional perfusion differences • relative GMR calculated from PET studies • functional pattern typical of AD by SPECT and PET used to calculate ratio of average perfusion or metabolism in affected areas divided by unaffected areas • ROC analysis performed	**General findings** • metabolism ratios in normals significantly related to age, and metabolic differences between normal and AD less obvious in old age • perfusion and metabolism ratios significantly lower in AD than in VD and controls • no significant differences between perfusion ratio and severity of dementia or age **ROC analysis** • for discrimination between AD and controls PET had marginally significant (p = .05) advantage (PET Se = 80% at Sp = 100%; SPECT Se = 80% at Sp = 65%) • PET false negatives all in marginally demented patients • SPECT false negatives scattered across range of dementia severity • for differential diagnosis of AD versus VD, PET was superior to SPECT **Conclusions** • both PET and SPECT can distinguish AD from controls • PET is superior in differentiating AD from VD **Study design limitations** • no definitive diagnosis in any DAT subjects

Abbreviations:
AD, Alzheimer's disease
CT, computed tomography
HMPAO, hexamethylpropylene amine oxime
NINCDS-ADRDA, National Institute of Neurological and Communicative Disorders and Stroke-Alzheimer's Disease and Related Disorders Association
NINDS-AIREN,
OM, orbitomeatal
ROC, receiver operating characteristic
ROI, region of interest
Se, sensitivity
Sp, specificity
SPECT, single photon emission tomography
VD, vascular dementia

Table 6 Study design models: Diagnostic thinking efficacy of alternatives to FDG PET for Alzheimer's disease

Notes: *The studies tabulated below were identified by searches of MEDLINE files for the years 1991 to 1996, using the terms "Alzheimer's disease" and "diagnosis".*

Van Gool, et al., analyzed two ongoing cohort studies in the Netherlands to estimate the diagnostic accuracy of SPECT in elderly patients presenting for an initial evaluation for mild dementia, and to determine the incremental contribution of SPECT after a thorough clinical and laboratory examination. However, this study did not follow patients to death and autopsy.

The Oxford Project to Investigate Memory and Aging (Jobst, et al.) used that project's cohorts of patients with dementia and age-matched controls; analyses are based on subjects who were followed to death and autopsy. These studies are classified as "diagnostic thinking efficacy" studies because they are presented in a form that allows estimation of the risk of AD for an individual using age-specific prevalence data and the likelihood ratio from that individual's CT.

Study	*Subjects/Methods*	*Results/Comments*
Van Gool, et al., 1995 *Academic Medical Center, Amsterdam, The Netherlands*	**Purpose** to conduct a study to address methodologic deficiencies of existing SPECT studies: • definition of incremental value of SPECT after careful clinical examination • spectrum bias in studies including relatively young patients and patients with advanced disease • diagnostic utility of SPECT in patients representing diagnostic challenge (mildly affected elderly) **Study design** cross sectional with controls **Cases** 110 patients > 65 year referred for first evaluation of dementia **Controls** 18 subjects recruited from prospective community-based study of mental functioning in elderly (65-85 years); controls had suboptimal cognitive scores **Methods** • initial clinical diagnosis using CAMDEX-N interview schedule, with specification of whether SPECT expected to contribute to diagnostic certainty by both individual neurologist and consensus panel • lab, CT, SPECT temporoparietal perfusion data then used with DSM III-R and NINCDS-ADRDA criteria for final diagnosis • all ancillary tests (lab and imaging) scored re contribution to change from initial to final diagnosis • diagnostic classification reconsidered after 6 months, FU to 2 years in some patients • SPECT images scored by consensus of 2 of 3 neurologists blinded to clinical findings (semiquantitative analysis) • Se of SPECT calculated for multiple perfusion value cut points • Sp of SPECT calculated using images from non-demented controls to avoid potential contribution of AD encephalopathy to other primary diagnoses • ROC analyses conducted • a priori requirement that a claim of substantial contribution of SPECT to diagnostic process would be validated if proportion of patients whose diagnosis changed ≤ 20%	**Final diagnoses** • 68 probable AD according to NINCDS-ADRDA criteria • 42 other (multi infarct dementia, unspecified dementia, mixed dementia) **Operating characteristics of SPECT** at temporoparietal perfusion cut point of 0.79 • non demented controls, Sp = 89% • all probable AD, Se = 43% • probable AD < 80 years, Se = 56% • probable AD > 80 years, Se = 29% **Contribution of SPECT to diagnosis** • before imaging, SPECT was expected to contribute to diagnostic certainty in 26% of patients • actual contribution of SPECT to 8% of final diagnoses (5 DAT, 3 mixed dementia, 1 rule out mixed dementia) **Authors' comments** • routine SPECT has limited value in evaluating elderly demented patients • disagreement with other published results attributed to selection bias in other studies (relatively young patients or those with advanced disease, or highly selected healthy controls)

Study	Subjects/Methods	Results/Comments
Jobst, et al., 1992a 1992b 1994 *Oxford University, UK*	**Purpose** to develop simple, clinically applicable diagnostic measures for AD **Study design** cohorts with and without DAT followed to autopsy **Cases** 45 cases; definitive postmortem diagnosis of AD (using pathologic criteria of Khachaturian) **Controls** •16 with other causes of dementia (documented by postmortem histopathology) • 8 with no dementia in life and no postmortem CNS pathology • 84 living without evidence of cognitive decline **General methods** subjects with and without dementia had detailed, repeated annual assessments (full neuropsychological, psychiatric, physical and radiological screening) until death, when autopsy was carried out **CT methods** quantitative evaluation of atrophy performed from temporal lobe-oriented CT acquired along long axis of medial temporal lobe: • linear measurement of narrowest thickness of medial temporal lobe on right or left side at level of brain stem between its anterior and posterior margins • interrater reliability tested by comparing measurements of 2 observers, blind to diagnosis, on scans from 127 subjects (mean difference of 0.15, good agreement) **SPECT methods** scans assessed semiquantitatively and by consensus • high interrater reliability • perfusion scans graded from 0 (no deficit) to 3 (severe deficit breaching cortical rim)	**CT** at cut point of < 0.79 multiple of median (< 5th percentile for age of controls) = AD (5% false positives): Se = 94%; Sp = 93.5% • measurement falls below fifth percentile for age in confirmed AD at least 4 years before death and certainly prior to onset of severe dementia • results permit estimation of risk for AD for individual using age-specific prevalence data and likelihood ratio from individual's value for minimum thickness of medial temporal lobe (LR = ratio of height of gaussian curve for AD to height of curve for controls at a patient's value of the minimum thickness of the temporal lobe) **SPECT** Grade ≥ 2 perfusion deficit in temporoparietal cortex = AD: Se = 96%; Sp = 89% **CT + SPECT** using < 5th percentile for age medial temporal lobe measurement on CT and grade ≥ 2 perfusion deficit in temporoparietal cortex on SPECT = AD (3% false positives): Se = 90%; Sp = 97% **Authors' conclusions** • cognitive evaluation plus CT plus SPECT decreased average false positive rate of 25% using clinical criteria alone to less than 5% • Main immediate value of findings is to identify groups at high enough risk of AD to justify recruitment into clinical trials of potential new therapies

Abbreviations:
AD, Alzheimer's disease
CT, computed tomography
Se, sensitivity
Sp, specificity
SPECT, single photon emission tomography

VI. REFERENCES: Background and diagnostic accuracy/diagnostic thinking efficacy studies

Arai H, Terajima M, Miura M, Higuchi S, Muramatsu T, Machida N, et al. Tau in cerebrospinal fluid: a potential diagnostic marker in Alzheimer's disease. *Annals of Neurology* 1995;38:649-52.

Azari NP, Pettigrew KD, Schapiro MB, Haxby JV, Grady CL, Pietrini P, et al. Early detection of Alzheimer's disease: a statistical approach using positron emission tomography data. *Journal of Cerebral Blood Flow and Metabolism* 1993;13:438-47.

Burdette JH, Minoshima S, Vander Borght T, Tran DD, Kuhl DE. Alzheimer disease: improved visual interpretation of PET images by using three-dimensional stereotaxic surface projections. Radiology 1996;198:837-43.

Crismon ML. Tacrine: first drug approved for Alzheimer's disease. *Annals of Pharmacotherapy* 1994;28:744-51.

Davis KL, Thal LJ, Gamzu ER, Davis C, Woolson RF, Gracon IS et al., and the Tacrine Collaborative Study Group: A double-blind, placebo-controlled multicenter study of tacrine for Alzheimer's disease. *New England Journal of Medicine* 1992;327:1253-9.

Fazekas F, Alavi A, Chawluk JB, Zimmerman RA, Hackney D, Bilaniuk L, et al. Comparison of CT, MR, and PET in Alzheimer's dementia and normal aging. *Journal of Nuclear Medicine* 1989; 1607-15.

Gearing M, Mirra SS, Hedreen JC, Sumi SM, Hansen LA, Heyman A. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part X. Neuropathology confirmation of the clinical diagnosis of Alzheimer's disease. *Neurology* 1995;45:461-6.

Geldmacher DS, Whitehouse PJ. Evaluation of dementia. *New England Journal of Medicine* 1996;335:330-6.

Giacometti AR, Davis PC, Alazraki NP, Malko JA. Anatomic and physiologic imaging of Alzheimer's disease. *Clinics in Geriatric Medicine* 1994;10:277-98.

Growdon JH. Treatment for Alzheimer's disease? (editorial) *New England Journal of Medicine* 1992;327:1306-8.

Herholz K, Perani D, Salmon E, Granck G, Fazio F, Heiss WD, et al. Comparability of FDG PET studies in probable Alzheimer's disease. *Journal of Nuclear Medicine* 1993;34:1460-6.

Hoffman JM, Hanson MW, Coleman RE. Clinical positron emission tomography imaging. *Radiology Clinics of North America* 1993;31:935-59.

Hoffman JM, Hanson MW, Welsh KA, Earl N, Paine S, Delong D, et al. Interpretation variability of ^{18}FDG-positron emission tomography studies in dementia. *Investigative Radiology* 1996;31:316-22.

Jarvik LF, Matsuyama SS, Ghui , Scheibel AB, Vinters HV. Autopsy diagnoses of Alzheimer disease: independent reviews and clinical implications. *International Journal of Geriatric Psychiatry* 1995;10:505-10.

Jobst KA, Smith AD, Barker CS, Wear A, King EM, Smith A, et al. Association of atrophy of the medial temporal lobe with reduced blood flow in the posterior parietotemporal cortex in patients with a clinical and pathological diagnosis of Alzheimer's disease. *Journal of Neurology, Neurosurgery and Psychiatry* 1992a;55:190-4.

Jobst KA, Smith AD, Szatmari M, Molyneux A, Esiri ME, King E, et al. Detection in life of confirmed Alzheimer's disease using a simple measurement of medial temporal lobe atrophy by computed tomography. *Lancet* 1992b; 340:1179-83.

Jobst KA, Hindley NJ, King E, Smith AD. The diagnosis of Alzheimer's disease: a question of image? *Journal of Clinical Psychiatry* 1994;55(11, suppl):22-31.

King EMF, Smith A, Jobst, K.A. *Age and Ageing* 1993;22:209-14.

Kippenhan JS, Barker WW, Nagel J, Grady C, Duara, R. Neural-network classification of normal and Alzheimer's disease subjects using high-resolution and low-resolution PET cameras. *Journal of Nuclear Medicine* 1994;35:7-15.

Kukull WA, Larson EB, Reifler BV, Lampe TH, Yerby MS, Hughes JP. The validity of 3 clinical diagnostic criteria for Alzheimer's disease. *Neurology* 1990;40:1364-9.

Larson EB, Kukull WA, Katzman RL. Cognitive impairment: dementia and Alzheimer's disease. *Annual Review of Public Health* 1992;13:431-49.

Links JM, Devous MD. Detection and comparison of patterns in images (editorial). *Journal of Nuclear Medicine* 1994;35:16-7.

Mann UM, Mohr E, Gearing M, Chase TN. Heterogeneity in Alzheimer's disease: progression rate segregated by distinct neuropsychological and cerebral metabolic profiles. *Journal of Neurology, Neurosurgery and Psychiatry* 1992;55:956-9.

Maisey M, Jeffery P. Clinical applications of positron emission tomography. *British Journal of Clinical Psychiatry* 1991;45.

McCormick WC, Larson EB. Dementia. In Panzer RJ, Black ER, Griner PF, eds.: *Diagnostic Strategies for Common Medical Problems*. American College of Physicians, Philadelphia 1991.
McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* 1984; 34:939-44.

Mielke R, Pietrzyk U, Jacobs A, Fink, GR, Ichimiya A, Kessler J, et al. HMPAO SPET and FDG PET in Alzheimer's disease and vascular dementia: comparison of perfusion and metabolic pattern. *European Journal of Nuclear Medicine* 1994;1053-60.

Morris JC. Differential diagnosis of Alzheimer's disease. *Clinical Geriatric Medicine* 1994;10:257-76.

Motter R, Vigo-Pelfrey C, Kholodenko D, Barbour R, Johnson-Wood K, Galasko D, et al. Reduction of ß-Amyloid Peptide$_{42}$ in the cerebrospinal fluid of patients with Alzheimer's disease. *Annals of Neurology* 1995;38:643-48.

Nalbantoglu J, Gilfix BM, Bertrand P, Robitaille R, Gauthier S, Rosenblatt DS, Poirier J. Predictive value of apolipoprotein E genotyping in Alzheimer's disease: results of an autopsy series and an analysis of several combined studies. *Annals of Neurology* 1994; 36:889-95. National Institute on Aging/Alzheimer's Association Working Group. Apolipoprotein E genotyping in Alzheimer's disease. *Lancet* 1996;347:1091-5.

Phelps CS, Hutson A. Estimating diagnostic test accuracy using a "fuzzy gold standard". *Medical Decision Making* 1995;15:44-57.

Post SG. Alzheimer's disease: ethics and the progression of dementia. *Clinical Geriatric Medicine* 1994;10:379-94.

Rapoport SI. Positron emission tomography in Alzheimer's disease in relation to disease pathogenesis: a critical review. *Cerebrovascular and Brain Metabolism Reviews* 1991;3:297-335.

Reiman EM, Caselli RJ, Yun LS, Chen K, Bandy D, Minoshima S, et al. Preclinical evidence of Alzheimer's disease in persons homozygous for the ε4 allele for apolipoprotein E. *New England Journal of Medicine* 1996;334:752-8.

Rocca WA. Frequency, distribution, and risk factors of Alzheimer's disease. *Nursing Clinics of North America* 1994;29:101-11.

Rockwood K, Stadnyk K. The prevalence of dementia in the elderly: a review. *Canadian Journal of Psychiatry* 1994;39:253-7.

Rogers SL, Friedhoff LT, Apter JT, Richter RW, Hartford JT, Walshe TM, et al. The efficacy and safety of donepezil in patients with Alzheimer's disease: results of a US multicentre, randomized, double-blind, placebo-controlled trial. *Dementia* 1996;7:293-303.

Salmon E, Sadzot B, Mazuet P, Degueldre C, Lemaire C, Rigo P, et al. Differential diagnosis of Alzheimer's disease with PET. *Journal of Nuclear Medicine* 1994;35:391-98.

Scinto LFM, Daffner KR, Dressler D, Ransil BI, Rentz D, Weintraub S, et al. A potential noninvasive neurobiological test for Alzheimer's disease. *Science* 1994;266:1051-4.

Siegel BV, Buchsbaum MS, Starr A, Mohs RC, Neto DC. Glucose metabolic rate and progression of illness in Alzheimer's disease. *International Journal of Geriatric Psychiatry* 1995;10:659-67.

Small GW, Mazziotta JC, Collins MT, Baxter LR, Phelps ME, Mandelkern MA, et al. Apolipoprotein E type 4 allele and cerebral glucose metabolism in relatives at risk for familial Alzheimer disease. *JAMA* 1995;273:942-47.

Van Gool WA, Walstra GJM, Teunisse S, Van der Zant FM, Weinstein HC, Van Royen EA. Diagnosing Alzheimer's disease in elderly, mildly demented patients: the impact of routine single photon emission computed tomography. *Journal of Neurology* 1995;242:401-5.

Wagstaff AJ, McTavish D. Tacrine: a review of its pharmacodynamic and pharmacokinetic properties, and therapeutic efficacy in Alzheimer's disease. *Drugs and Aging* 1994;4:510-40.

Whitehouse PJ, Geldmacher DS. Pharmacotherapy for Alzheimer's disease. *Clinical Geriatric Medicine* 1994;10:339-50.

VII. REFERENCES: Technical efficacy studies

Alavi A, Newberg AB, Souder E, Berlin JA. Quantitative analysis of PET and MRI data in normal aging and Alzheimer's disease: atrophy weighted total brain metabolism and absolute whole brain metabolism as reliable discriminators. *Journal of Nuclear Medicine* 1993;34:1681-7.

DeCarli D, Grady CL, Clark CM, Katz DA, Brady DR, Murphy DGM, et al. Comparison of positron emission tomography, cognition, and brain volume in Alzheimer's disease with and without severe abnormalities of white matter. *Journal of Neurology, Neurosurgery, and Psychiatry* 1996;60:158-67.

Duara R, Grady C, Haxby J, Sundaram M, Cutler NR, Heston L, et al. Positron emission tomography in Alzheimer's disease. *Neurology* 1986;36:879-87.

Ferris SH, de Leon MJ, Wolf AP, George AE, Reisberg B, Christman DR, et al. Positron emission tomography in dementia. *Advances in Neurology* 1983;38:123-9.

Foster NL, Chase TN, Mansi L, Brooks R, Fedio P, Patronas NJ, et al. Cortical abnormalities in Alzheimer's disease. *Annals of Neurology* 1984;16:649-54.

Guze BH, Hoffman JM, Mazziotta JC, Baxter LR, Phelps ME. Positron emission tomography and familial Alzheimer's disease: a pilot study. *Journal of the American Geriatric Society* 1992;40:120-3.

Haxby JV, Grady CL, Duara R, Schlageter N, Berg G, Rapoport SI. Neocortical metabolic abnormalities preceded nonmemory cognitive defects in early Alzheimer's-type dementia. *Archives of Neurology* 1986;43:882-5.

Kennedy AM, Frackowiak RSJ, Newman SK, Bloomfield PM, Seaward J, Roques P, et al. Deficits in cerebral glucose metabolism demonstrated by positron emission tomography in individuals at risk of familial Alzheimer's disease. *Neuroscience Letters* 1995;186:17-20.

Kumar A, Schapiro MB, Grady C, Haxby JV, Wagner E, Salerno JA, et al. High-resolution PET studies in Alzheimer's disease. *Neuropsychopharmacology* 1991;4:35-46.

de Leon MJ, Ferris SH, George AE, Reisberg B, Christman DR, Kricheff II,et al. Computed tomography and positron emission transaxial tomography evaluations of normal aging and Alzheimer's disease. *Journal of Cerebral Blood Flow and Metabolism* 1983;3:391-4.

McGeer EG, Peppard RP, McGeer PL, Tuokko H, Crockett D, Parks R, et al. ^{18}Fluorodeoxyglucose positron emission tomography studies in presumed Alzheimer cases, including 13 serial scans. *Canadian Journal of Neurologic Science* 1990;17:1-11.

Melzer CC, Zubieta JK, Brandt J, Tune LE, Mayberg HS, Frost JJ. Regional hypometabolism in ALzheimer's disease as measured by positron emission tomography after correction for effects of partial volume averaging. *Neurology* 1996;47:454-61.

Minoshima S, Frey KA, Foster NL, Kuhl DE. Preserved pontine glucose metabolism in Alzheimer disease: a reference region for functional brain image (PET) analysis. *Journal of Computer Assisted Tomography* 1995;19:541-7.

Minoshima S, Frey KA, Koeppe RA, Foster NL, Kuhl DE. A diagnostic approach in Alzheimer's disease using three-dimensional stereotactic surface projections of fluorine-18-FDG PET. *Journal of Nuclear Medicine* 1995;36:1238-48.

Slansky I, Herholz K, Pietrzyk U, Kessler J, Grond M, Mielke R, et al. Cognitive impairment in Alzheimer's disease correlates with ventricular width and atrophy-corrected cortical glucose metabolism. *Neuroradiology* 1995;37:270-7.

Small GW, Okonek A, Mandelkern MA, La Rue A, Chang L, Khonsary A, et al. Age-associated memory loss: initial neuropsychological and cerebral metabolic findings of a longitudinal study. *International Psychogeriatrics* 1994;6:23-44.

VIII. REFERENCES: Excluded studies

Exclusion criteria were:

- number of DAT cases < 12
- duplicated or superseded by subsequent study from the same institution
- behavioral or cognitive activation rather than resting FDG PET
- radiopharmaceutical other than FDG
- case series (without controls)
- diagnostic accuracy efficacy study where PET data were interpreted visually but blinding was not noted
- DAT diagnostic criteria not specified
- insufficient information to judge comparability of case and control groups, details of imaging protocol, whether visual or quantitative analysis of PET data used, or type of PET data analysis used

Azari NP, Pietrini P. Preclinical stages in subjects at risk for neurological disorders: can PET-FDG tell us more? *Journal of Neurology* 1995;242:112-4.

Benson DF, Kuhl DE, Phelps ME, Cummings JL, Tsai SY. Positron emission computed tomography in the diagnosis of dementia. *Trans American Neurology Association* 1981;106:68-71.

Benson DR, Kuhl DE, Hawkins RA, Phelps ME, Cummings JL, Tsai SY. The fluorodeoxyglucose ^{18}F scan in Alzheimer's disease and multi-infarct dementia. *Archives of Neurology* 1983;40:711-4.

Besson JAO, Crawford JR, Evans NT, Gemmell HG, Roeda D. PET imaging in Alzheimer's disease. *Journal of the Royal Society of Medicine* 1992:231-34.

Chase TN, Foster NL, Fedio P, Brooks R, Mansi L, Di Chiro G. Regional cortical dysfunction in Alzheimer's disease as determined by positron emission tomography. *Annals of Neurology* 1984; 15(suppl):S170-4.

Chase TN, Foster NL, Fedio P, Mansi L, Brooks R, Kessler R, et al. Cognitive and cerebral metabolic function in early and advanced Alzheimer's disease. *Monographs in Neural Sciences* 1984;11:176-9.

Chawluk JB, Dann R, Alavi A, Hurtig HI, Gur RE, Resnick S, et al. The effect of focal cerebral atrophy in positron emission tomographic studies of aging and dementia. *Nuclear Medicine and Biology* 1990;17:797-804.

Cutler NR, Haxby JV, Duara R, Grady CL, Kay AD, Kessler RM, et al. Clinical history, brain metabolism, and neuropsychological function in Alzheimer's disease. *Annals of Neurology* 1985; 18:298-309.

Duara R, Barker W, Loewenstein D, Pascal S, Bowen B. Sensitivity and specificity of positron emission tomography and magnetic resonance imaging studies in Alzheimer's disease and multi-infarct dementia. *European Neurology* 1989;29(suppl 3):9-15.

Duara R, Barker WW, Chang J, Yoshii F, Loewenstein DA, Pascal S. Viability of neocortical function shown in behavioral activation state PET studies in Alzheimer disease. *Journal of Cerebral Blood Flow and Metabolism* 1992;12:927-34.

Farkas T, Ferris SH, Wold AP, de Leon MJ, Christman DR, Reisberg B, et al. ^{18}F-2-deoxy-2-fluoro-D-glucose as a tracer in the positron emission tomographic study of senile dementia. *American Journal of Psychiatry* 1982;139:352-3.

Ferris SH, de Leon MJ, Wolf AP, Farkas T, Christman DR, Reisberg B, et al. Positron emission tomography in the study of aging and senile dementia. *Neurobiology of Aging* 1980;1:127-31.

Foster NL, Chase TN, Fedio P, Patronas NJ, Brooks RA, Di Chiro, G. Alzheimer's disease: focal cortical changes shown by positron emission tomography. *Neurology* 1983;33:961-5.

Friedland RP, Budinger TF, Ganz E, Yano Y, Mathis CA, Koss B, et al. Regional cerebral metabolic alterations in dementia of the Alzheimer type: positron emission tomography with [^{18}F]fluorodeoxyglucose. *Journal of Computer Assisted Tomography* 1983;7:590-8.

Friedland RP, Budinger TF, Brant-Zawadzki M, Jagust WJ. The diagnosis of Alzheimer-type dementia: a preliminary comparison of positron emission tomography and proton magnetic resonance. *JAMA* 1984;252:2750-2.

Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Medical Decision Making* 1991; 11:88-94.

Fukuyana H, Ogawa M, Yamaguchi H, Yamaguchi S, Kimura J, Yonekura Y, et al. Altered cerebral energy metabolism in Alzheimer's disease: a PET study. *Journal of Nuclear Medicine* 1994;35:1-6.

Haxby JV, Duara R, Grady CL, Cutler NR, Rapoport SI. Relations between neuropsychological and cerebral metabolic asymmetries in early Alzheimer's disease. *Journal of Cerebral Blood Flow and Metabolism* 1985;5:193-200.

Herholz K, Adams R, Kessler J, Szelies B, Grond M, Heiss WD. Criteria for the diagnosis of Alzheimer's disease with positron emission tomography. *Dementia* 1990;1:156-64.

Jagust WJ, Friedland RP, Budinger TF. Positron emission tomography with [^{18}F]fluorodeoxyglucose differentiates normal pressure hydrocephalus from Alzheimer-type dementia. *N Neurol* 1985;48:1091-6.

Jagust WJ, Friedland RP, Budinger TF, Koss E, Ober B. Longitudinal studies of regional cerebral metabolism in Alzheimer's disease. *Neurology* 1988;38:909-12.

Kennedy AM, Rossor MN, Frackowiak RSJ. Positron emission tomography in familial Alzheimer disease. *Alzheimer Disease and Associated Disorders* 1995;9:17-20.

Kessler J, Herholz K, Grond M, Heiss WD. Impaired metabolic activation in Alzheimer's disease: a PET study during continuous visual recognition. *Neuropsychologica* 1991;29:229-43.

Kippenhan JS, Barker WW, Pascal S, Nagel J, Duara R. Evaluation of a neural-network classifier for PET scans of normal and Alzheimer's disease subjects. *Journal of Nuclear Medicine* 1992; 33:1459-67.

Kuhl DE, Metter EJ, Riege WH. Patterns of cerebral glucose utilization in depression, multiple infarct dementia, and Alzheimer's disease. *Brain Imaging and Brain Function* 1985:211-6.

Kuhl DE, Metter EJ, Riege WH, Hawkins RA, Mazziotta JC, Phelps ME, et al. Local cerebral glucose utilization in elderly patients with depression, multiple infarct dementia, and Alzheimer's disease. *Journal of Cerebral Blood Flow and Metabolism* 1983;3(Supplement 1):S494-5.

McGeer PL, Kamo H, Harrop R, McGeer EG, Martin WRW, Pate BD,et al. Comparison of PET, MRI, and CT with pathology in a proven case of Alzheimers disease. *Neurology* 1986;36:1569-74.

Mentis MJ, Weinstein EA, Horwitz B, McIntosh AR, Pietrini P, Alexander GE, et al. Abnormal brain glucose metabolism in the delusional misidentification syndromes: a positron emission tomography study in Alzheimer disease. *Biologic Psychiatry* 1995;38:438-49.

Powers WJ, Perlmutter JS, Videen TO, Herscovitch P, Griffeth LK, Royal HD, et al. Blinded clinical evaluation of positron emission tomography for diagnosis of probable Alzheimers disease. *Neurology* 1992;42:765-70.

Salmon E, Franck G. Positron emission tomographic study in Alzheimer's disease and Pick's disease. *Archives of Gerontology and Geriatrics* 1989;Suppl 1:241-7.

Smith GS, de Leon MJ, George AE, Kluger A, Volkow ND, McRae T, et al. Topography of cross-sectional and longitudinal glucose metabolic deficits in Alzheimer's disease. *Archives of Neurology* 1992;49:1142-50.

Swerdlow R, Marcus DL, Landman J, Kooby D, Frey W, Freedman ML. Brain glucose metabolism in Alzheimer's disease. *American Journal of Medical Science* 1993;308:141-4.

Szelies B, Mielke T, Herholz K, Heiss WD. Quantitative topographical EEG compared to FDG PET for classification of vascular and degenerative dementia. *Electroencephalography and Clinical Neurophysiology* 1994;91:131-9.

Valladares-Neto DC, Buchsbaum MS, Evans WJ, Nguyen D, Nguyen P, Siegel BV, et al. EEG delta, positron emission tomography, and memory deficit in Alzheimer's disease. *Neuropsychobiology* 1995;31:173-81.

Appendix 9

Experience With PET in VHA

Author: Elizabeth Adams, R.R.T., M.P.H., *Management & Program Analyst,*
Technology Assessment Program

Appendix 9

Experience With PET in VHA

Veterans Health Administration (VHA), shares the ownership and operation of 10 positron emission tomography (PET) imaging facilities with some of its academic affiliates. Significant resource commitments are associated with the acquisition, maintenance, and operation of these facilities.

In late 1993, the Acting Under Secretary for Health in VHA requested that the Management Decision and Research Center (within Health Services Research and Development Service) conduct a rigorous examination of the agency's investment in PET. The Acting Under Secretary asked two questions:

- *Should the VHA add more PET Centers?*
- *How is PET used in VHA today?*

The Advisory Committee to the PET assessment focused the assessment on the use of PET in diagnosing diseases relevant to the veteran population and on collecting information about PET imaging utilization, center operations, and clinical and research activities.

To obtain information on the experience with PET within VHA, a written survey was distributed prior to the site visits, and a follow-up survey was sent out in December, 1995. Site visits were conducted by a MDRC Technology Assessment Management and Program Analyst and an external consultant from August through October, 1994.

This text briefly summarizes the information obtained by the MDRC Technology Assessment Program on the experience at 11 VHA PET centers.

I. BACKGROUND

PET is a relatively new addition to the repertoire of clinical diagnostic tests available both within and outside VHA. All but three of the VHA PET facilities became operational after 1990, and the information collected through the site visits and surveys represents preliminary data on VHA experience with the technology.

Of the 12 initially approved PET sites, 11 were fully operational at the time of the assessment; support for the twelfth had been withdrawn. After completion of the site visits, support for another PET center was discontinued by local VA medical center administration. At the time of release of this report, 10 VHA PET centers were in operation. Locations of the VHA PET centers are depicted in Figure 1 at the end of this section.

II. METHODS

A written survey addressing characteristics and staffing of PET installations, characteristics of the medical centers where the PET facilities were housed, and the types and volume of PET studies was distributed to each PET center approximately three to four weeks prior to the site visits.

Preparation for the site visits was made with the assistance of the PET director and/or the Chief, Nuclear Medicine Service at each VHA site, who acted as the primary contact person. The contact person was responsible for compiling a list of interview subjects and coordinating the interview schedule. For this interview schedule he or she was asked to include referring and non-referring physicians from all sharing partners and within four major specialties: cardiology, neurology, oncology, and psychiatry. These specialties represent the clinical areas where PET is most likely to be used. The contact persons were encouraged to include other specialties deemed important to the activities of their PET centers.

At nine sites, interviews were conducted over two days. The two remaining sites required only one day to cover their referral bases. Most interviews were completed in 30 minutes, and confidentiality of interview content was stressed.

The written survey and interview questionnaires may be found at the end of this appendix.

III. RESULTS

The information in this section was obtained from pre-site visit survey materials, from responses of interview subjects based on the interview questionnaires, and from observations made by the site visit team. Results of the pre-site visit survey are summarized in Tables 1 through 11 and are described in the sections “Characteristics of interview subjects,” “Characteristics of PET centers” and “Types and volumes of PET studies.” The section on “Costs” is also based on pre-site visit survey data. Results from the 1995 follow up survey are presented in Table 12.

Results of the site visit interviews are summarized in Tables 13 through 16 and are described in the sections “Barriers and incentives to PET use,” “Sharing agreements,” and “Research activity at VHA PET Centers.” Issues related to the negotiation and content of the sharing agreements and to the research activity of these PET centers were felt to be of sufficient importance to be discussed in separate sections. A summary is provided at the end of this section.

A. Characteristics of interview subjects

The composition of interview subjects is presented in the following pages in Table 1 and is summarized in Tables 2 through 4. There was an equitable distribution of clinical specialties represented among interview subjects. The majority of subjects interviewed were classified as referrers, of which 7% referred fewer than 5 patients annually and another 34% referred an unknown number of patients annually for PET scans. The vast majority of interview subjects had multiple job roles consisting primarily of clinical and research duties with some administrative component, reflecting the academic environments in which these PET centers were placed.

B. Characteristics of PET centers

Table 5 compares the ancillary services available at VHA PET centers. To the extent that these services might be associated with the use of PET (e.g., a wide range of cardiology or neurology diagnostic services are available), most of the VHA centers seem to have an appropriate and relatively equivalent array of services.

Table 6 provides general information on the characteristics of VHA PET facilities. Table 7 summarizes the data in Table 6. Most of the PET centers became operational within the last three years. The data reflect a range of scanner models used across sites. Ownership and location of the scanner were evenly distributed among VAMCs and their sharing partners, whereas ownership and location of the cyclotron tended to be concentrated among the sharing partners (i.e., academic affiliates). Thus, the sharing partner was inclined to be the primary source of the radiotracers used in PET scanning.

All sites used cyclotron produced radioisotopes as tracers. Fluorodeoxyglucose (FDG) was the only radiopharmaceutical common to all sites. Many sites generated and used ^{15}O-water and ^{13}N-ammonia, as well. Responsibility for personnel was evenly divided among VAMCs and their sharing partners.

Table 1: Site Visit Interview Subjects According to Specialty, Job Role, and Referral Status

Note: Data reporting annual referral patterns of clinicians and researchers excluded interview subjects classified as administrators only and non-referring specialties.

Site	Specialty	Interview Subjects With a Single Role			Interview Subjects With Multiple Roles				Total (% Site Total)	Annual Referral Patterns of Clinicians and Researchers Listed in Columns to Left			
		Administrator	Clinician	Researcher	Administrator/ Clinician	Administrator/ Researcher	Clinician/ Researcher	Admin/ Clinician/ Researcher		Non-referrer	Referrer (number of patients referred annually)		
											(1-5)	(>5)	Unknown
A	nonclinical	1							1 (5)				
	cardiology						3	1	4 (21)				4
	neurology						3	2	5 (26)			1	3
	oncology					1	1		2 (10)	1			1
	psychiatry			1		1	2		4 (21)	1		1	2
	other	1						2	3 (16)				2
							Total for Site		**19 (100)**				
B	non-clinical	1							1 (6)				
	cardiology						1	3	4 (25)			4	
	neurology				1		2		3 (19)		2	1	
	oncology						2	2	4 (25)	1		3	
	psychiatry							2	2 (13)	2			
	other	1						1	2 (13)				
							Total for Site		**16 (100)**				
C	non-clinical	3				1			4 (31)	1			
	cardiology							2	2 (15)		1	1	
	neurology							2	2 (15)			2	
	oncology							1	1 (8)	1			
	psychiatry							1	1 (8)	1			
	other							3	3 (23)	1		1	
							Total for Site		**13 (100)**				

Site	Specialty	Interview Subjects With a Single Role			Interview Subjects With Multiple Roles				Total (% Site Total)	Annual Referral Patterns of Clinicians and Researchers Listed in Columns to Left			
		Administrator	Clinician	Researcher	Administrator/ Clinician	Administrator/ Researcher	Clinician/ Researcher	Administrator/ Clinician/ Researcher		Non-referrer	Referrer (number of patients referred annually)		
											(1-5)	(>5)	Unknown
D	non-clinical	3							3 (17)				
	cardiology						1	1	2 (11)			2	
	neurology						3	1	4 (22)	1	1	2	
	oncology							3	3 (17)	1		2	
	psychiatry							2	2 (11)	2			
	other						1	3	4 (22)	1		2	
							Total for Site		**18 (100)**				
E	non-clinical			1					1 (6)	1			
	cardiology						3	1	4 (22)			4	
	neurology				1		1	2	4 (22)	1	2		
	oncology		1					2	3 (17)	1		2	
	psychiatry							3	3 (17)	2			1
	other		1				1	1	3 (17)	1	1		
							Total for Site		**18 (100)**				
F	non-clinical	1							1 (7)				
	cardiology						2	1	3 (21)		2		
	neurology						4		4 (29)			2	2
	oncology		1				1	1	3 (21)			3	
	psychiatry							1	1 (7)	1			
	other		1					1	2 (14)				
							Total for Site		**14 (100)**				

Site	Specialty	Interview Subjects With a Single Role			Interview Subjects With Multiple Roles				Total (% Site Total)	Annual Referral Patterns of Clinicians and Researchers Listed in Columns to Left			
		Administrator	Clinician	Researcher	Administrator/ Clinician	Administrator/ Researcher	Clinician/ Researcher	Administrator/ Clinician/ Researcher		Non-referrer	Referrer (number of patients referred annually)		
											(1-5)	(>5)	Unknown
G	non-clinical	1							1 (8)				
	cardiology						1	1	2 (15)	2			
	neurology						2	1	3 (23)	1		1	1
	oncology						1		1 (8)	1			
	psychiatry						1		1 (8)				1
	other						2	3	5 (38)	2			
							Total for Site		**13 (100)**				
H	non-clinical	5							5 (15)				
	cardiology					4			4 (12)				4
	neurology						6		6 (18)				6
	oncology						4	1	5 (15)				5
	psychiatry						6	1	7 (21)				6
	other	1					3	2	6 (18)				4
							Total for Site		**33 (100)**				
I	non-clinical								0 (0)				
	cardiology						2	1	3 (16)			1	2
	neurology						2	4	6 (32)	2		3	1
	oncology		1				1		2 (11)	1			1
	psychiatry						2	1	3 (16)		1	2	
	other				1		2	2	5 (26)			1	1
							Total for Site		**19 (100)**				

Site	*Specialty*	*Interview Subjects With a Single Role*			*Interview Subjects With Multiple Roles*				*Total (% Site Total)*	*Annual Referral Patterns of Clinicians and Researchers Listed in Columns to Left*			
		Administrator	*Clinician*	*Researcher*	*Administrator/ Clinician*	*Administrator/ Researcher*	*Clinician/ Researcher*	*Administrator/ Clinician/ Researcher*		*Non-referrer*	*Referrer (number of patients referred annually)*		
											(1-5)	*(>5)*	*Unknown*
J	non-clinical								0 (0)				
	cardiology						2	1	3 (23)			3	
	neurology						2	2	4 (31)	1		1	2
	oncology						2	1	3 (23)			1	1
	psychiatry						1		1 (8)			1	
	other	1						1	2 (15)				
							Total for Site		**13 (100)**				
K	non-clinical	1							1 (6)				
	cardiology							2	2 (13)			2	
	neurology						1	1	2 (13)			2	
	oncology						1	2	3 (19)			2	1
	psychiatry						1	1	2 (13)	2			
	other						1	5	6 (38)	1		1	
							Total for Site		**16 (100)**				

Table 2: Summary of Site Visit Interview Subjects According to Specialty

Specialty	*Total (% Total Subjects Interviewed)*
non-clinical	18 (9)
cardiology	33 (17)
neurology	43 (22)
oncology	30 (16)
psychiatry	27 (14)
other	41 (21)
TOTAL	192 (100)

Table 3: Summary of Site Visit Interview Subjects According to Referral Patterns

Annual Referral Patterns	*Total (% Total Subjects Interviewed)**
non-referrer	34 (23)
1-5 patients	10 (7)
>5 patients	54 (36)
referred unknown number	51 (34)
TOTAL	149 (100)

*Note: Total number excludes interview subjects classified as non-clinical and non-referring specialties

Table 4: Summary of Site Visit Interview Subjects According to Job Role

Job Role (s)	*Administrator*	*Clinician*	*Researcher*	*Administrator/ Clinician*	*Administrator/ Researcher*	*Clinician/ Researcher*	*Administrator/ Clinician/ Researcher*	*TOTAL*
Total (% Total Subjects Interviewed)	20 (10)	5 (3)	2 (1)	3 (2)	7 (4)	77 (40)	78 (41)	192 (100)

Table 5: A Comparison Of Ancillary Services Offered At Each VHA PET Site

Service	***Sites Offering Service (%)***
Alcohol Dependency Treatment Unit	100
Cancer Center	82
Cardiac Cath Lab	100
Cardiac ICU	100
Cardiac Surgery Program	100
Electron Microscopy	73**
Epilepsy Program	100
Geriatric Research Education & Clinical Center (GRECC)	91**
Health Psychology Program	45**
Hemodialysis In-Center Care	91
Home Dialysis and CAPD Training	100
Hypertension Screening and Treatment Program	100
Medical ICU	100
Mental Hygiene Clinic	100
Neuropsychological Testing	100
Nursing Home Care Unit	91
Patient Health Education Program	91
Prosthetic and Sensory Aid Service	100
PTSD Program	91
Pulmonary Function Lab	100
Sickle Cell Screening Program	53
Speech Pathology Lab	91**
Surgical ICU	100
Women’s Health Center	64**
Other: (Nuclear Medicine Network)	9

** reflects uncertainty of some respondents in whether a service was offered; actual percentage may be higher

Table 6: General Information of VHA PET Sites as of Fiscal Year 1994

Site	*Start-up Year*	*Scanner Model*	*Owner of Scanner*	*Owner of Cyclotron*	*Owner of Radiochem Lab*	*Location of Camera*	*Location of Cyclotron and Lab*	*Personnel Employer*	*FDG Source*	*Radiopharmaceuticals Generated or Used*
A	1993	Positron Posicam	VA and SP	VA and SP	VA and SP	VA	VA	VA and SP	VA	^{18}F-FDG, ^{13}N-Ammonia, ^{18}F-DOPA
B	1992	Siemens 951/31	VA and SP	SP	SP	SP	SP	SP	SP	^{18}F-FDG, ^{13}N-Ammonia
C	1992	Siemens 951R	VA and SP	SP	SP	SP	SP	SP	SP	^{18}F-FDG, ^{13}N-Ammonia, ^{15}O-Water
D	1988	Siemens 931/08-12	VA and SP	SP	SP	SP	SP	SP	SP	^{18}F-FDG, ^{13}N-Ammonia, ^{15}O-Water, ^{11}C-Acetate
E	1993	Siemens 951R	VA and SP	VA and SP	VA and SP	SP	SP	SP	SP	^{18}F-FDG, ^{13}N-Ammonia, ^{15}O-Water, ^{11}C-Acetate
F	1979	Siemens 933 & GE Advance	VA and SP	SP	SP	VA	SP	VA and SP	SP	^{18}F-FDG, ^{13}N-Ammonia, ^{15}O-Water, ^{18}F-DOPA, ^{18}F-Methane, ^{18}F-Lomafloxacin, ^{62}Cu-PTSM, ^{60}Cu-PTSM, ^{94m}Tc-Teboroxyine, ^{94m}Tc-Sestamibi
G	1992	Siemens 953B	VA	VA	VA	VA	VA	VA	VA	^{18}F-FDG, ^{15}O-Water
H	1992	GE 4096	VA and SP	VA and SP	VA and SP	SP	SP	VA and SP	SP	^{18}F-FDG, ^{13}N-Ammonia, ^{15}O-Water
I	1985	Siemens 953/31	VA	VA	VA	VA	VA	VA	VA	^{18}F-FDG, ^{13}N-Ammonia, ^{15}O-Water
J	1991	Siemens 931/04	VA	None	None	VA	Private Source	VA	Private Source	^{18}F-FDG, ^{18}F
K	1993	Siemens 951/31	VA and SP	VA and SP	SP	VA	SP	VA and SP	SP	^{18}F-FDG, ^{13}N-Ammonia, ^{15}O-water

SP=Sharing Partner
^{11}C= carbon-11
^{60}Cu=copper-60
^{62}Cu= copper-62
DOPA= dihydroxyphenylalanine
^{18}F= fluorine-18
FDG= fluorodeoxyglucose
^{13}N= nitrogen-13
^{15}O= oxygen-15
PTSM= pyruvaldehyde bis(N^4-methylthiosemicorbazone)
^{94m}Tc=Technetium-94m

Table 7: Summary of the General Characteristics of the VHA PET Sites

Characteristic *Description*	*Options*	***Frequency*** **Number (%)**
Start up year	1979	1 (9)
	1985	1 (9)
	1988	1 (9)
	1991	1 (9)
	1992	4 (36)
	1993	3 (27)
Scanner model (some sites have > 1 scanner)	Positron Posicam	1 (8)
	Siemens 951/31	3 (25)
	Siemens 951/R	2 (17)
	Siemens 931/08-12	1 (8)
	Siemens 933	1 (8)
	GE advance	1 (8)
	Siemens 953B	1 (8)
	GE 4096	1 (8)
	Siemens 931/04	1 (8)
Owner of scanner	VA	3 (27)
	VA and sharing partner	8 (73)
Owner of cyclotron	VA	2 (20)
	Sharing partner	4 (40)
	VA and sharing partner	4 (40)
Owner of radiochemistry lab	VA	2 (20)
	Sharing partner	5 (50)
	VA and sharing partner	3 (30)
Location of camera	VA	6 (55)
	Sharing partner	5 (45)
Location of cyclotron	VA	3 (27)
	Sharing partner	7 (64)
	Private source used	1 (9)
Personnel employer	VA	3 (27)
	Sharing partner	4 (36)
	VA and sharing partner	4 (36)
FDG source	VA	3 (27)
	Sharing partner	7 (64)
	Private vendor	1 (9)
Radiopharmaceuticals	FDG	11 (100)
	^{13}N-ammonia	9 (82)
	F-DOPA	2 (18)
	^{15}O-water	8 (73)
	other	4 (36)

C. Types and volumes of PET studies

Tables 8 and 9 present information on the types and volumes of clinical and research studies conducted at each VHA PET center and its academic affiliate for Fiscal Year 1994; Tables 10 and 11 provide the same information for Fiscal Year 1993. Table 12 presents data from a follow-up survey on total patient volume for Fiscal Year 1995 and related issues.

Inter-site comparisons using these data were problematic for a number of reasons. There was significant variability among protocols with respect to scan time and resources used; some patients were scanned multiple times. Most sites logged their utilization according to patient and protocol, rather than the actual time involved in acquiring PET studies. Variations in PET technology across sites also affected utilization, as the scanning process took longer with older models.

Volume comparisons across sites using total number of scans would require a standardized workload unit and prospective data collection. The MDRC Technology Assessment (TA) Program felt that expressing patient volume according to the number of patients studied best reflected the referral base of each site. Therefore, comparisons using total number of patients rather than total number of scans were made.

The MDRC TA Program was asked to evaluate the level of patient activity at each site. Therefore, animal studies were not included in the volume data. Four of the eleven sites performed PET scans on animals, for a total of 279 studies in 1994 and 256 in 1993.

The tables indicate that a wide range of types and volumes of studies are performed across the VHA system. In 1993, there was a small overall disparity in utilization between VHA and its academic affiliates (45% and 55% of the total studies performed, respectively). The majority of PET studies was conducted for clinical purposes in neurology applications, followed by cardiology and oncology. The vast majority of research activity was in neurology and psychiatry.

In 1994, the disparity in utilization between VHA and its academic affiliates had widened to 31% and 69%, respectively. Clinical neurology applications continue to be the main focus of activity at these PET centers, followed by oncology and cardiology. The vast majority of research activity was in neurology and psychiatry with a growing interest in oncology.

Data on total volume for Fiscal Year 1995 were obtained from all but one site, which is no longer supported by VHA, but continues to be supported by the university affiliate. In 1995 the disparity in utilization between VA and non-VA studies was decreased to 41% and 59%, respectively. Seven sites reported an increased demand for clinical PET studies, while one reported a decrease and two reported no change. The increase in clinical interest was attributed largely to clinical oncology applications. Two sites expressed an increased use of PET in psychiatric and neurologic research.

Table 8: Patient Volume at VHA PET Sites for Fiscal Year 1994

Note: Definition of clinical oncology studies varied across sites. Clinical psychiatry studies listed were for the diagnosis of manic depression or schizophrenia. Those data not reported or available were indicated as "N/A".

Site	Patient Type	Cardiology Studies					Neurology Studies					Psychiatry Studies		Oncology Studies		Other	Clinical Total	Research Total	Site Total
		Clinical				Research	Clinical				Research	Clinical	Research	Clinical	Research	Research	(% Site Total)	(% Site Total)	(% of system-wide total)
		Viability	Ischemic Heart Disease	Other	Subtotal		Epilepsy	Tumor vs. necrosis	Other	Subtotal									
C	VA	N/A	N/A	0	0	0	3	0	0	3	0	0	0	0	0	0	3 (1)	0 (0)	339 (17)
	non-VA	N/A	N/A	0	0	0	3	15	0	18	267	0	0	0	51	0	18 (5)	318 (94)	
E	VA	26	1	0	27	0	6	0	0	6	0	0	0	21	0	0	54 (17)	0 (0)	327 (16)
	non-VA	48	18	0	66	0	74	2	5	81	1	1	0	122	2	0	270 (83)	3 (1)	
I	VA	7	0	0	7	10	51	0	1	52	10	0	0	54	0	0	113 (41)	20 (7)	278 (14)
	non-VA	0	0	0	0	0	16	0	0	16	0	0	0	129	0	0	145 (52)	0 (0)	
G	VA	6	0	0	6	0	0	0	0	0	21	0	15	0	0	0	6 (3)	36 (17)	208 (11)
	non-VA	0	0	0	0	0	4	1	0	5	113	0	48	0	0	0	5 (2)	161 (77)	
A	VA	N/A	N/A	0	0	0	0	0	48	48	25	0	0	0	0	0	48 (25)	25 (13)	193 (10)
	non-VA	N/A	N/A	0	0	0	N/A	N/A	77	77	43	0	0	0	0	0	77 (40)	43 (22)	
F	VA	0	0	0	0	0	15	0	5	20	0	0	0	0	1	0	20 (12)	1 (0.1)	171 (9)
	non-VA	5	22	0	27	14	13	23	0	36	22	0	41	0	10	0	63 (37)	87 (51)	
D	VA	14	0	N/A	14	0	45	0	0	45	52	0	14	0	0	25	59 (39)	91 (61)	150 (8)
	non-VA	N/A	N/A	N/A	0	N/A	N/A	N/A	N/A	0	N/A	N/A	N/A	N/A	N/A	0	0 (0)	0 (0)	
B	VA	2	2	0	4	0	15	12	3	30	0	3	0	20	0	0	57 (54)	0 (0)	105 (5)
	non-VA	8	7	0	15	0	5	6	1	12	0	1	0	20	0	0	48 (46)	0 (0)	
K	VA	23	3	0	26	0	0	3	0	3	9	0	0	0	14	0	29 (34)	23 (26)	87 (4)
	non-VA	8	0	0	8	0	5	0	0	5	16	0	0	0	6	0	13 (15)	22 (25)	
H	VA	2	0	0	2	0	0	0	0	0	0	0	17	0	0	0	2 (2)	17 (20)	84 (4)
	non-VA	5	1	0	6	2	9	0	0	9	33	0	15	0	0	0	15 (18)	50 (59)	
J	VA	13	0	0	13	0	0	1	1	2	0	0	0	7	0	0	22 (44)	0 (0)	50 (2)
	non-VA	4	0	0	4	0	17	4	0	21	0	0	0	3	0	0	28 (56)	0 (0)	
TOTAL		171	54	0	225	26	281	67	141	489	612	5	150	376	84	25	1095	897	1992 (100

Table 9: A Comparison of VA to Non-VA Patient Volume Within Each Clinical and Research Application Across All VHA PET Sites for Fiscal Year 1994

Note: Definition of clinical oncology studies varied across all sites. Clinical psychiatry studies listed were for the diagnosis of manic depression or schizophrenia. Those data not reported or available were indicated as "N/A".

PATIENT TYPE	CARDIOLOGY VOLUME (% COLUMN TOTAL)				NEUROLOGY VOLUME (% COLUMN TOTAL)				PSYCHIATRY VOLUME (% COLUMN TOTAL)		ONCOLOGY VOLUME (% COLUMN TOTAL)		OTHER VOLUME (% COLUMN TOTAL)	TOTAL (% COLUMN TOTAL)
	Clinical Studies				Clinical Studies									
	Viability	Ischemic Heart Disease	Other	Research	Epilepsy	Tumor vs. Necrosis	Other	Research	Clinical	Research	Clinical	Research	Research	
VA	93 (54)	6 (11)	0	10 (38)	135 (48)	16 (24)	58 (41)	108 (18)	3 (60)	46 (31)	108 (29)	15 (18)	25 (100)	623 (31)
non-VA	78 (46)	48 (89)	0	16 (62)	146 (52)	51 (76)	83 (59)	504 (82)	2 (40)	104 (69)	268 (71)	69 (82)	0 (0)	1369 (69)
Total VA + non-VA	171 (100)	54 (100)	0 (100)	26 (100)	281 (100)	67 (100)	141 (100)	612 (100)	5 (100)	164 (100)	376 (100)	84 (100)	25 (100)	1992 (100)
Total (% PET activity systemwide)	171 (9)	54 (3)	0 (0)	26 (1)	281 (14)	67 (3)	141 (7)	612 (31)	5 (0.2)	150 (8)	376 (19)	84 (4)	25 (1)	1992 (100)

Table 10: Patient Volume at VHA PET Sites for Fiscal Year 1993

Note: Definition of clinical oncology studies varied across sites. Clinical psychiatry studies listed were for the diagnosis of manic depression or schizophrenia. Those data not reported or available were indicated as "N/A".

Site	Patient Type	Cardiology Studies					Neurology Studies					Psychiatry Studies		Oncology Studies		Other	Clinical Total	Research Total	Site Total
		Clinical				Research	Clinical				Research	Clinical	Research	Clinical	Research	Research	(% Site Total)	(% Site Total)	(% of system-wide total)
		Viability	Ischemic heart disease	Other	Subtotal		Epilepsy	Tumor vs. necrosis	Other	Subtotal									
I	VA	4	0	0	4	13	115	0	21	136	71	0	1	162	0	0	302 (69)	85 (19)	439 (25)
	non-VA	0	0	0	0	0	16	11	0	27	1	0	0	24	0	0	51 (12)	1 (0.2)	
A	VA	58	58	0	116	0	0	0	48	48	31	0	0	0	0	0	164 (59)	31 (11)	278 (16)
	non-VA	4	4	0	8	0	20	5	0	25	50	0	0	0	0	0	33 (12)	50 (18)	
D	VA	18	18	3	39	0	0	0	51	51	31	0	31	0	0	26	90 (51)	88 (49)	178 (10)
	non-VA	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A			
E	VA	13	1	0	14	0	0	6	0	6	0	0	0	8	0	0	28 (19)	0 (0)	149 (9)
	non-VA	35	1	0	36	0	33	6	0	39	0	0	0	46	0	0	121 (81)	0 (0)	
F	VA	7	0	0	7	0	8	2	0	10	0	0	0	0	0	0	17 (12)	0 (0)	140 (8)
	non-VA	8	10	0	18	3	35	5	0	40	27	0	33	0	0	2	58 (42)	65 (46)	
C	VA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 (0)	0 (0)	137 (8)
	non-VA	0	0	0	0	0	0	0	0	0	3	0	134	0	0	0	0 (0)	137 (100)	
G	VA	0	0	0	0	0	0	1	1	2	28	0	0	0	0	0	2 (2)	28 (28)	99 (6)
	non-VA	0	0	0	0	0	12	4	0	16	45	0	8	0	0	0	16 (16)	53 (54)	
J	VA	10	0	0	10	0	0	2	10	12	0	0	0	6	0	1	28 (31)	1 (1)	91 (5)
	non-VA	22	0	0	22	0	21	4	1	26	0	0	13	1	0	0	49 (54)	13 (14)	
H	VA	0	0	0	0	0	0	0	0	0	0	0	27	0	0	0	0 (0)	27 (34)	80 (5)
	non-VA	0	0	0	0	0	2	0	0	2	18	0	25	0	0	8	2 (2)	51 (64)	
B	VA	4	4	0	8	0	4	7	2	13	0	0	0	5	0	0	26 (39)	0 (0)	66 (4)
	non-VA	10	10	0	20	0	6	2	1	9	0	0	0	11	0	0	40 (61)	0 (0)	
K	VA	33	0	0	33	0	0	0	0	0	1	0	0	0	3	0	33 (52)	4 (6)	63 (4)
	non-VA	3	0	0	3	0	12	11	0	23	0	0	0	0	0	0	26 (41)	0 (0)	
TOTAL		229	106	3	338	16	284	66	135	485	306	0	272	263	3	37	1086	634	1720 (100)

Table 11: A Comparison of VA to Non-VA Patient Volume Within Each Clinical and Research Application Across All VHA PET Sites for Fiscal Year 1993

Note: Definition of clinical oncology studies varied across sites. Clinical psychiatry studies listed were for the diagnosis of manic depression or schizophrenia.

PATIENT TYPE	CARDIOLOGY VOLUME (% OF ALL CARDIOLOGY STUDIES)				NEUROLOGY VOLUME (% OF ALL NEUROLOGY STUDIES)				PSYCHIATRY VOLUME (% OF ALL PSYCHIATRY STUDIES)		ONCOLOGY VOLUME (% OF ALL ONCOLOGY STUDIES)		OTHER VOLUME (% OF ALL OTHER STUDIES)	TOTAL (% TOTAL STUDIES)
	Clinical Studies				Clinical Studies									
	Viability	Ischemic Heart Disease	Other	Research	Epilepsy	Tumor vs. Necrosis	Other	Research	Clinical	Research	Clinical	Research	Research	
VA	147 (64)	81 (76)	3 (100)	13 (81)	127 (45)	18 (27)	133 (99)	162 (53)	0	59 (22)	181 (69)	3 (100)	27 (73)	954 (55)
non-VA	82 (36)	25 (24)	0	3 (19)	157 (55)	48 (73)	2 (1)	144 (47)	0	213 (78)	82 (31)	0	10 (27)	766 (45)
Total VA + non-VA	229 (100)	106 (100)	3 (100)	16 (100)	284(100)	66 (100)	135 (100)	306 (100)	0	272 (100)	263 (100)	3 (100)	37 (100)	1720 (100)
Total (% PET activity systemwide)	229 (13)	106 (6)	3 (0.1)	16 (1)	284 (17)	66 (4)	135 (8)	306 (18)	0	272 (16)	263 (15)	3 (0.2)	37 (2)	1720 (100)

Table 12: Follow-up Survey of Activity at VHA PET Sites for Fiscal Year 1995

Note: Data are presented in order of total patients scanned.

A list of abbreviations is located at the end of the table.

Site	Number of Patients Studied		Number of Animals Studied	Change in demand for clinical studies from FY '94 to FY '95 and comments	Impact of proposed changes in FDA regulations on center	Comments on proposed changes in FDA regulations	Comments on trends experienced in last year
	VA	non-VA					
B	204	416	0	**increased** due to: • interest in oncology • growing interest among referring physicians	will modify PET center operations in FY '96	anticipate the need to submit NDA or ANDA in FY '96	• low interest in cardiac applications • 90-95% of studies in clinical oncology to determine extent of disease and response to therapy where traditional anatomic studies (CT,MRI) are equivocal • approaching daily capacity of system at 3-4 cases/day
I	301	230	0	**increased** due to: • interest in oncology	will modify PET center operations in FY '96	refurbishing lab area to include an automated FDG system	increasing interest and utilization of ^{15}O-water studies with applications in various central nervous system activation paradigms and vascular studies of extremities
A	241	210	3	**increased** due to: • enhanced contact with referral staff • educating medical community • reimbursement negotiated with local carrier • approval by local authority for oncology procedures	will modify PET center operations in FY '96	will work to meet requirements for good manufacturing practices to meet FDA regulations	• increased acceptance of PET by medical community • plan to increase marketing efforts • will further develop support for oncology referrals and reimbursements
G		357	2	**no change**	modified PET center operations in FY '95	District Counsel has written an opinion stating that FDA has no authority over PET scanning at VAMC	
C	3	312	44	**decreased** due to: • lack of reimbursement (for mostly non-VA patients) • fewer cardiology but more oncology studies	• modified PET center operations in FY '95 and FY '96 • plan to coordinate activities with other PET centers	• changes would be problematic should site wish to manufacture pharmaceuticals for distribution • state-of-the-art facility; some requirements already in place • confusion regarding new GMPs and their application to PET, given that all products are already tested before administration to patient	
E	77	183	72	**no known changes**	will modify PET center operations in FY '96	plan to obtain Investigative New Drug application	
K	199*	35*	11*	**increased** due to: • interest in oncology	none	none reported	none reported

Site	Number of Patients Studied		Number of Animals Studied	Change in demand for clinical studies from FY '94 to FY '95 and comments	Impact of proposed changes in FDA regulations on center	Comments on proposed changes in FDA regulations	Comments on trends experienced in last year
	VA	non-VA					
J	99	41	0	**increased** due to: • installation of whole body scanner • interest in oncology	none	none reported	• PET has replaced CT as next test following x-ray in evaluation of solitary pulmonary nodules • PET is used clinically in patients with lung cancer, colorectal cancer, lymphoma, and melanoma
H	25	99	0	**increased** due to: • expanding referral base (for evaluation of patients with seizure disorders and tumor patients for pre-operative brain mapping • reimbursement from private carriers and limited considerations from Medicare and Medicaid for seizure studies	will modify PET center operations in FY '96	establishing a contractual agreement with an outside company to use the VHA PET facility for distribution of FDG; agreement will include using the FDG provided by them under their NDA for all clinical studies	• plan to increase clinical emphasis as a source of revenue and to continue this trend in FY 1996 by implementing PET oncology studies • PET center staffing is decreasing; currently one PET technologist at the center; one sharing partner discontinued support for personnel.
D	104	0	0	**increased** due to: • interest among referring physicians to use PET for differential diagnosis and designing treatment plans	will modify PET center operations in FY '96	GMP regulations may jeopardize present sharing agreements	• majority of referrals are for tumor localization in clinical oncology • PET used for research in alcohol, alcohol treatment, and PTSD • PET used to study medical/social problems; findings announced in media
F**	---	---	---	---	---	---	---
Total (% Total)	1251 (40%)	1863 (60%)	---	---	---	---	---

*excluding a total of 77 research studies
**VHA discontinued its support in 1995; now supported by university affiliate

Abbreviations:
GMP=good manufacturing practices
CT=computerized tomography
MRI= magnetic resonance imaging
PTSD=post traumatic stress disorder
NDA=new drug application
ANDA=abbreviated new drug application

D. Costs

The data for this section were not tabulated because of variations in the definitions of some cost elements across sites and among sharing partners. The major costs at each PET site were: equipment amortization; maintenance contracts for the scanner; maintenance contracts for the cyclotron; scanner-related supplies; cyclotron supplies including target materials; and personnel, particularly highly skilled radiochemists, clinical and research specialists, analysts and programmers. Other significant costs included installation and maintenance of pneumatic tube systems used to transport radioactive isotopes between facilities, and start-up funding to cover the overhead costs for the initial years of operations.

In an effort to offset these costs, some sites generated revenue by selling cyclotron products to private PET facilities, while others extended their catchment area to include a broader patient base. At one site the decision was made to maintain low operating costs by purchasing cyclotron products from a private source, rather than producing its own. However, this limited its research capabilities.

One site recommended that, to offset the high and often unexpected maintenance costs of the scanner and cyclotron, an escrow account be established from equal contributions made by the sharing partners. A “roving” maintenance team supported by VHA to service all VHA PET centers was suggested as another potential solution. The disadvantage of this solution is that since technical expertise in PET is limited, there is a considerable likelihood that these technicians would be subsequently recruited by the private sector.

E. Barriers and incentives to PET use

Table 13 lists the barriers and incentives to PET use that were discussed in the interviews, and that may contribute to the range of frequencies seen in Tables 8-12. Statements that appear to conflict reflect the diverse opinions and interests of the interview subjects. Table 14 lists the recommendations mentioned during the interviews for improving the management of PET centers and increasing utilization of PET. In addition, a number of VHA PET centers provided examples of processes with which they had addressed some of these issues; these processes are listed as “best practices” in Table 15. Information regarding issues related to the FDA and trends in utilization may also be found in Table 12.

1. *General issues*- The site visit interviews indicated that there are significant organizational, professional, scientific, and reimbursement issues yet to be resolved before PET becomes more widely diffused. Ambiguities in the interpretation of FDA regulations regarding the use of FDG and other radiotracers for clinical purposes contributed to variations in the authority (federal versus state) under which PET sites chose to govern their operations, and subsequently, in the types of clinical and research PET studies conducted (See Tables 12 and 13). Proposed changes in FDA regulations related to manufacturing practices of radiopharmaceuticals will likely result in modification of operations at most sites. Generic PET issues such as limited FDA approved clinical PET applications and lack of demonstrated clinical utility were felt to perpetuate the perception of the general medical community and regulators that PET is primarily a research tool. These issues were also believed to contribute to inconsistent reimbursement policies.

Sites that obtained reimbursement for clinical studies generally developed *a priori* consensus-building efforts among payers and providers within their communities in exchange for data collection. Sites less successful in obtaining reimbursement often encountered external organizational and political obstructions which prevented PET scans from being performed. Clinicians described the pre-approval process, typically conducted on a case-by-case basis by many private payers, as untimely and impractical for many clinical needs. At one site local politics impeded reimbursement; in an effort to counter strong union pressure for comprehensive coverage and lower costs to their members, the state commercial payers applied rigorous coverage exclusion criteria to technologies they classified as “experimental.”

VHA contributed significantly to overall PET activity by committing substantial resources toward the initial start-up of twelve PET centers systemwide. Although support for one site was discontinued, the remaining eleven sites continued to receive VHA support for subsequent sharing agreements (at the time of the site visit). Centralized strategic planning involving the distribution, construction and maintenance of these centers was seen as necessary to the overall investment into costly high technologies such as PET. Nevertheless, these processes were described as frustrating, inefficient, and protracted.

PET center operations were thought to be adversely affected by the lack of vision and commitment in Headquarters. For example, one VAMC received funding to purchase PET technology, but not the provisions with which to house it. Likewise, new technologies and their associated services often required support for operations beyond the acquisition year to cover staffing needs and revenue shortfalls in the early start-up years before the centers became fully operational. Funding for replacement parts and maintenance were usually not included in the initial acquisition arrangement. Many administrators expressed concerns of having to support new programs and services with existing funding levels.

Variations in VHA’s financial commitment to the PET centers appeared to be related to the degree to which this support was continued by local medical center administrators. The degree of local support was reflected in the content of the sharing agreements, in the commitment to house the PET center and assume its high overhead costs, and in the administrators’ tolerance of their centers’ financial losses.

Although the PET centers’ main mission or focus ranged from a primarily research to a primarily clinical orientation, most sites acknowledged that a mix of clinical and research activity was desirable. Some administrators viewed PET as an expensive but valuable technology for furthering research, expanding clinical services, and enhancing prestige, and were willing to accept some financial losses. An example of this was the willingness of some VAMC administrators to use patient care dollars to finance a PET center with a predominantly research mission. Other PET centers were asked to cut costs. One hospital director threatened to eliminate the PET center, because it had failed to sustain itself financially. (Since the site visit, the university affiliate assumed total financial support of the PET center. The VAMC continues to lease space to them for PET operations.)

2. *Practical considerations*- Interview subjects cited several practical considerations that contributed to volume shortfalls (See Table 13). The technical characteristics of the PET camera affected scanning time, which may have taken from 1.5 hours to several hours. Newer models scanned faster and produced higher quality images. Another frequently noted problem was the availability of radioactive tracers. Their production and use were often timed to coincide with other scheduled studies in an effort to minimize costs, but in doing so, scheduling and access to the scanner may have been affected. Many clinicians expressed the need for more staff education on the clinical applications of PET, although they also acknowledged that its clinical utility needed further study.

Inadequate staffing (particularly radiochemists) was cited as impeding the conduct of certain studies. In VA hospitals, PET centers' hours of operation were frequently curtailed by inflexible tours of duty, restrictions in overtime salary, and restrictions and/or cutbacks in the number of Full Time Equivalent Employees. The ability to conduct PET studies with complex radioactive tracers, whose development is very time- and resource-intensive, is contingent upon the availability and qualifications of its radiochemist. Four PET centers cited the need for a qualified radiochemist as a major influence on the variety and volume of patient studies. The supply of radiochemists in the general PET community is limited; competition for these specialists is intense.

3. *Ratio of VA to non-VA patients*- Several issues contributed to differences in the ratio of VA to non-VA patients studied across sites. The location of the PET center and issues related to patient transport were noted. Difficulties obtaining reimbursement for patient transport to the PET center were cited by the affiliates as an important barrier to access if the center was at the VAMC, whereas transport for VA patients was fully covered by most VAMCs. Location played an important role in determining which patients could be scanned, as patients too medically unstable to be transported were unable to be scanned. Problems specific to VA and to VA patients included poor patient compliance in keeping scheduled appointments and the perception by private sector patients that the quality of care delivered by VHA was substandard, or that the availability of services provided by VA hospitals was restricted to VA patients only. Many PET center directors expressed frustration at not having the authority or resources to properly market their services to the private sector.

Although reimbursement for VA patients' clinical studies was more consistent, the widening disparity of VA to non-VA patients studied in Fiscal Year 1994 (See Tables 8 and 9) indicate that other factors may influence veterans' access to PET. At many sites, VA investigators expressed concern for the lack of available research funding, especially within VHA. The inability to attract VA patients for PET scans may reflect either a lower burden of illness among veterans with respect to the general population, or the disparity between the underlying characteristics of veteran patients, who are frequently more debilitated, and a protocol's inclusion criteria. Moreover, a clinical PET study may not be requested by a referring physician if the test is not felt to contribute information that would increase diagnostic certainty and affect subsequent choice of treatment.

4. *Competition*- Competition at many levels affected the degree to which PET was used. The site visit team observed competition among clinical specialties for access to PET, between PET and other technologies, and among PET centers in the same city. At three sites the use of administrative and regulatory mechanisms to impede some investigators' access to PET was mentioned. At one site, where the PET center was located at the affiliate, an ineffective sharing agreement permitted little recourse on the part of the VAMC with which to gain and maintain equal access; fostered the perception among clinicians that VA administrators valued preserving the relationship between the sharing partners over the interests of the VAMC and its veterans; and allowed for an imbalance in representation of both specialties and sharing partners on the PET Oversight Committee. The site visit team identified one site that developed a process to overcome these barriers based on a model at the National Institutes of Health to facilitate the review and approval of their PET protocols, thereby assuring access to PET for all (See Table 15).

Competing interests from other functional imaging technologies including Single Photon Emission Computed Tomography (SPECT) and functional Magnetic Resonance Imaging (MRI) tended to dilute both administrative and academic support for PET. At several sites, clinicians were more likely to favor SPECT over PET because of wider acceptance among clinicians and third party payers and greater accessibility to clinicians. Competition with other active PET centers within the community may have affected both the referral base from which private patients were recruited and the ability to attract both PET specialists and scarce research funding.

5. *Referral base*- Strong academic and clinical interests in functional imaging were important incentives for supporting technologies such as PET. The depth and breadth of the clinical and research referral base at each site influenced the types of applications studied, the kinds of patients included in these studies, and the proportion of clinical and research studies conducted. Gaps in selected areas of patient volume were often reflections of low or nonexistent interest of some specialties in PET. The site visit team discovered that neither the depth nor the breadth of the referral base extended far. Intensive recruitment efforts of specialists interested in PET took months in many cases. At one site, the loss of one epilepsy specialist effectively eliminated any activity in that area until recruitment efforts were completed.

Data from Fiscal Year 1995 (See Table 12) suggest a growing interest among referrers, particularly in clinical oncology. This may be due, in part, to the results of educational and marketing efforts made by PET center staff in recent years and to the growing body of PET literature reflecting interest in clinical oncology applications.

6. *Sharing agreements*- The sharing agreement process was cited as reflecting the trust and respect between the partners. All negotiators mentioned the need to balance the relationship between the sharing partners and to protect their individual interests. The negotiating team typically included representatives from Fiscal Service and the Director's Office. The degree to which the Director's Office participated in these negotiations varied across sites; the most active participation produced some of the most successful arrangements (See Table 15). To comply with VA policy (regarding recent changes in policy memo #2) PET directors with dual appointments were excluded from the negotiations. Consequently, the negotiations could not benefit from the insight of the individual who was most familiar with the needs of the service.

Three centers had no sharing agreements with their academic affiliates. In one case the relationship between sharing partners was strained, and in another, the PET center was located at and totally supported by the VAMC. The third center was also located at and supported by the VAMC, but had sharing agreements with local providers and individual researchers, rather than with its academic affiliate.

In these sharing agreements, PET center cost sharing varied from an even distribution between partners to covering partial costs. Those VAMCs sharing the cost burden with academic affiliates typically used estimated volume projections to compute the unit cost needed to meet overhead costs. In all cases, these volume projections were overestimated relative to actual experience. For those VAMCs with payback schedules based on these projections, payback had not been achieved, and the discrepancy was reconciled in other ways. Some sites renegotiated sharing agreements based on more realistic volume projections. One site developed a workload unit to better reflect true utilization of resources (See Table 15).

Some VA administrators expressed concern that the planning and subsequent construction costs of their affiliate's PET centers were extravagant, and that the VAMC was given minimal or no opportunity to participate in the planning. In recognition of this problem and to ensure that the VAMC would not be charged for excessive overhead costs, some sites successfully negotiated a lower patient charge based on an estimated cost or charge equal to the national average (as determined in a survey by the Institute for Clinical PET of PET centers in the U.S.).

Reconciliation of costs for VAMCs that had PET centers located at the academic affiliate were typically handled on a fee-for-service basis via a monthly billing system; the VAMC was charged for PET scans at a reduced rate compared to that of the private sector. Similarly, VAMCs charged for scans conducted on private patients by billing the academic affiliate, which then collected from private payers. The MDRC Technology Assessment Program found that the sharing arrangement most favorable for VAMCs with PET centers located at the academic affiliate was one that allowed for the full payment up front by the affiliate for its portion of the scanner. If contributing toward overhead costs, the VAMC was subsequently billed on a fee-for-service basis at a charge equal to or less than the national average. Another arrangement favorable to VAMCs was one in which a fixed number of "free" scans for VA patients was determined up front, in exchange for partial use of the scanner by other sharing partners. These arrangements insure that each VAMC recovers its portion of the investment up front, without risk of financial loss, should the volume projections be unfulfilled.

7. *Research activity at VHA PET centers*- Table 16 lists the wide range of research protocols available throughout VHA PET centers. Most research activity was in neurology and psychiatry, and to a lesser degree, in cardiology. There is a rapidly growing interest in oncology. Researchers in neurology and psychiatry view PET as critical for the progression of basic research in these areas. Those sites with existing funded research projects were more likely to sustain their research activity by attracting additional research funding and recruiting high level specialists. Most sites have a core infrastructure of nuclear medicine staff with specialized academic interests in functional imaging. One PET center is run by a neurologist with extensive funding in basic science neurology research and who has received substantial administrative support. Variations in research activity across sites reflect the degree to which the academic interests in functional imaging extended into other academic specialties.

PET research studies are generally more complex and quantitative than PET clinical studies. To sustain research activity, the following support was found to be important: state-of-the-art information systems, personnel to operate these systems, software, data analysts, a cyclotron within close proximity to produce the radioisotopes (including those with short half-lives) needed for most research studies, and, as discussed previously, a qualified radiochemist. At all sites, the reputation and expertise of the PET director and core PET center staff contributed positively to the willingness of medical staff and researchers to use PET as a clinical and research tool.

Several sites were found to be conducting similar research, yet many researchers were unaware of past or ongoing activity at other centers. Many researchers (particularly those in psychiatry) noted that they would like to coordinate research activity among sites, to make the most efficient use of available research funding. However, conflicts with interstate use of an Investigative New Drug protocol and the desire of some researchers to work independently affected their ability to cooperate with others.

Table 13: Results of Site Visit Interviews Reflecting Major Barriers and Incentives to the Use of PET Within Each Site

Note: Some comments may appear more than once within each site.

SITE		FACTORS AFFECTING THE USE OF PET			
		TYPE OF APPLICATIONS	VA/NON-VA PATIENTS	CLINICAL/RESEARCH STUDIES	OVERALL
A	Barriers to Use	• *reimbursement limited by restrictive criteria established by state authorities* • *limited approved use of FDG by FDA* • *greater degree of familiarity with and access to SPECT* • *low interest in clinical oncology among SPs*	• *difficulties obtaining reimbursement for ambulance transport* • *location:* - *not favorable for study of medically unstable non-VA patients* - *VA's reputation by private sector for poor quality care* - *perceived by some as only available to VA patients* • *inadequate coordination of services at VA between ER staff and testing labs, plus long wait to obtain most imaging tests make some research protocols not feasible for VA patients* • *conflicts between oncology clinic director and VA administration contribute to low clinical interest in PET* • *low interest in clinical oncology at affiliate*	• *investigators have access to many functional imaging modalities, diluting interest in solely PET* • *PET still viewed by many as a research tool only*	• *pressure by VA administration to become more cost-efficient; need to reduce FTEE of an existing skeletal staff* • *center's ability to develop and become self-sufficient felt to be impeded by lack of vision and commitment within CO* • *PET must compete with other imaging modalities for the support and interest of VAMC administration* • *more parking needed for outpatients* • *difficulties marketing PET services to community*
	Incentives to Use	• *significant level of funded research, particularly in psychiatry and neurology* • *PET viewed by the psychiatric community as critical for the progression of their research*	• *VA patients' scanning costs covered by VAMC* • *location of scanner more favorable to VA patients*	• *access of investigators to many functional imaging modalities, including PET*	• *strong academic interest in functional imaging* • *well integrated clinical staff among SPs*

SP= Sharing Partner

SITE		FACTORS AFFECTING THE USE OF PET			
		TYPE OF APPLICATIONS	VA/NON-VA PATIENTS	CLINICAL/RESEARCH STUDIES	OVERALL
B	Barriers to Use	• *lengthy start up time of ^{15}O-water production needed for neurology research affected credibility of PET center* • *SPECT more readily available to neurologists and psychiatrists* • *PET's clinical utility not demonstrated*	• *location; PET not accessible to unstable VA patients* • *lack of demonstrated clinical utility and restricted FDA approved clinical uses limits reimbursement for non-VA patients* • *non-VA patients covered by managed care require pre-approval, which may take several weeks; not always clinically practical* • *low clinical interest in cardiac PET studies at VA because of limited section budget and clinical utility* • *low interest in brain neuroimaging at affiliate*	• *no research funding obtained*	• *competition with another local active PET center* • *PET's clinical utility not clear:* - *limited approved use of FDG by FDA* - *limited clinical PET expertise among general staff* • *lengthy scan time related to PET's technical characteristics limits patient volume*
	Incentives to Use	• *growing interest in oncology PET applications* • *well-balanced interest of PET director in all PET applications* • *interest in brain neuroimaging in VA at GRECC*	• *VAMC reimburses clinical PET scans* • *interest in brain imaging in VA at GRECC*	• *well-balanced interest of PET director in all PET applications*	• *good working relationship among SPs* • *PET's role in tertiary care academic "centers of excellence" viewed favorably by VAMC administration*
C	Barriers to Use	• *no VA specialists in psychiatry and epilepsy surgery who use PET diagnostically* • *PET's low resolution in evaluating brain tumors* • *competition with MR, CT, SPECT for most clinical needs in psychiatry and neurology* • *inconsistent reimbursement from private sector*	• *no VA specialists in psychiatry and epilepsy surgery who use PET diagnostically* • *difficulties transporting VA patients affected by local topography* • *inadequate parking at VAMC* • *grant funding needed for VA researchers*	• *clinical utility not demonstrated* • *inconsistent reimbursement from private sector*	• *more physician education felt to be needed in PET technology* • *high costs; low reimbursement* • *PET's role in managed care not clear because clinical utility not demonstrated*
	Incentives to Use	• *increasing cardiology interest at VA*	• *increasing cardiology interest at VA* • *VA scans paid for on a fee-for-service basis*	• *research focus of PET center* • *important recruitment tool for attracting high level researchers*	• *well-respected PET center staff* • *PET viewed as important for institutional prestige*

SP=Sharing Partner

SITE		FACTORS AFFECTING THE USE OF PET			
		TYPE OF APPLICATIONS	VA/NON-VA PATIENTS	CLINICAL/RESEARCH STUDIES	OVERALL
D	Barriers to Use	• *perceived preferential scheduling of neurology studies* • *cardiology researchers perceive access to PET obstructed by the PET Operating Committee* • *competition with CT and MR in oncology work up* • *greater access to SPECT for cardiology studies*	• *location of scanner disadvantageous for VA patients* • *little reimbursement for clinical studies performed on non-VA patients*	• *research output affected by limited capabilities of radiochemist* • *little reimbursement for clinical studies for non-VA patients*	• *tension among SPs created by protracted sharing agreement process*
	Incentives to Use	• *expertise of PET director in neurologic applications*	• *reimbursement by VAMC for clinical PET scans on VA patients*	• *strong academic interest in functional imaging*	• *collaboration among SP investigators, especially in neurology and psychiatry* • *reputation of both SPs for delivering high quality patient care*
E	Barriers to Use	• *interest in neurologic PET applications not fostered by affiliate's Department Chair* • *oncologists focused in research areas other than PET*	• *inconsistent reimbursement for non-VA patients*	• *clinical utility not demonstrated* • *no specialist in clinical PET*	• *clinical utility in most areas not demonstrated* • *staff education needed on PET's clinical role* • *difficulties transporting VA patients transport to center*
	Incentives to Use	• *reimbursement for cardiac viability, epilepsy and brain tumors covered in non-VA patients* • *strong academic interest in functional imaging in psychiatry* • *growing interest in clinical oncology applications*	• *favorable reimbursement for VA patients*	• *equal emphasis on clinical and research use by PET director* • *grant-supported research interest in functional imaging*	• *VA considered an important player in the medical complex* • *PET viewed as a powerful recruitment tool* • *VAMC administration supportive of PET* • *clinical research data available to both SPs via development of shared computer archiving system*

SP= Sharing Partner

SITE		FACTORS AFFECTING THE USE OF PET			
		TYPE OF APPLICATIONS	VA/NON-VA PATIENTS	CLINICAL/RESEARCH STUDIES	OVERALL
F	Barriers to Use	• *easier access to and reimbursement for SPECT* • *pre-approval from private insurers required by IRB; impractical for many clinical applications*	• *lack of interest in PET among VA researchers:* - *low funding* - *little administrative support* • *clinical PET studies for VA patients not reimbursed by VAMC*	• *clinical PET studies for VA patients not reimbursed by VAMC* • *center deficient in radiochemistry expertise needed for research studies* • *distance of cyclotron from PET center; unable to scan using short-lived tracers* • *insufficient staff support for research project*	• *little reimbursement; clinical utility not demonstrated* • *VAMC administration unwilling to support PET with patient care dollars* • *strained relationship between SPs at administrative level* • *no nuclear medicine service at VAMC*
	Incentives to Use	• *well-integrated medical staff* • *reimbursement for myocardial viability studies* • *strong academic interest in neurology and cardiology imaging*	• *reputation of VAMC attracts large catchment area; able to recruit patients easily for cardiology and neurology studies*	• *reputation of VAMC as a quality institution* • *funded research interests in cardiology and neurology* • *PET's superior resolution over other modalities useful in some psychiatric research*	• *SPs connected via hallways:* - *facilitates ease of patient transport* - *fosters collegial relationship among clinical staff* • *highly regarded PET director* • *PET useful for recruiting medical staff*
G	Barriers to Use	• *weaker academic ties in cardiology and oncology* • *scanner equipped for brain studies only* • *IRB approval unsuccessful for most cardiology clinical research studies*	• *location of scanner not favorable to non-VA patients*	• *all PET studies require IRB approval; no clinical studies conducted which do not comply with FDA regulations* • *coordination of some research initiatives inhibited by physical distance between SPs*	• *total patient volume limited by finite hours of operation and tours of duty* • *mandatory IRB and RDRC approval of all protocols are perceived as obstruction mechanisms by some*
	Incentives to Use	• *PET director highly regarded for neurology PET expertise* • *strong academic ties in neurology and psychiatry*	• *location of scanner favorable to VA patients* • *new facility and reputation facilitates recruitment of non-VA patients*	• *center supported by funding for neurology PET research*	• *only PET scanner in the state* • *new facility;fosters positive reputation* • *PET center supported by VAMC administration*

SP=Sharing Partner

SITE		FACTORS AFFECTING THE USE OF PET			
		TYPE OF APPLICATIONS	VA/NON-VA PATIENTS	CLNICAL/RESEARCH STUDIES	OVERALL
H	Barriers to Use	• *clinical utility not demonstrated* • *limited FDA approved clinical use of FDG* • *little or no reimbursement for clinical studies*	• *financial hardship on center due to limitations imposed by sharing agreement* • *little funding for VA-sponsored research*	• *limited FDA approved clinical PET applications except for pre-surgical evaluation of epilepsy*	• *location of scanner; all patients require transportation to center* • *staffing limitations; a second radiochemist is needed* - *state mandates presence of radiochemist during cyclotron operations* - *neither SP is able to absorb costs of additional staffing* • *high operating costs*
	Incentives to Use	• *PET core staff expertise in neurologic and psychiatric PET applications* • *cardiology PET specialist recently recruited to staff* • *increasing interest in oncology applications*	• *mostly non-VA research funding obtained for PET studies* • *slow, but growing interest among VA researchers*	• *lengthy but well-coordinated protocol approval process* • *center supported largely by research grants*	• *strong reputation of both SPs within community* • *many successful sharing agreements negotiated between SPs* • *close integration of SPs at many levels* • *highly skilled personnel recruited as core staff* • *very cooperative and congenial atmosphere*
I	Barriers to Use	• *low cardiac research or clinical interest in PET* • *^{15}O-water not available for some neurology studies*	• *center not easily accessible to patients; located in back of hospital*	• *center designed more for clinical, rather than research purposes:* - *full-time radiochemist needed to develop complex tracers* - *PET support staff not always perceived as cooperative* • *location of radiochemistry lab; next to machine shop* • *availability of isotope affects expediency of inpatient testing*	• *daily patient volume restricted by limited operating hours* • *low morale among PET support staff*
	Incentives to Use	• *liberal private sector reimbursement for clinical oncology studies* • *strong academic interest in neurology and psychiatry PET applications*	• *VA patient scanning reimbursed by VAMC* • *location of scanner more favorable for VA patients*	• *reimbursement available for most clinical studies* • *large production capacity of cyclotron* • *VA PET center more accessible to investigators than university*	• *reputation of VAMC* • *PET director respected and liked* • *PET center generates revenue for VAMC; several sharing agreements negotiated with private sector*

SP=Sharing Partner

SITE		FACTORS AFFECTING THE USE OF PET			
		TYPE OF APPLICATIONS	VA/NON-VA PATIENTS	CLINICAL/RESEARCH STUDIES	OVERALL
J	Barriers to Use	• *perceived limited role in cardiac surgery work up* • *inconsistent reimbursement*	• *delays in approval by third party insurers for scanning non-VA patients*	• *no cyclotron or support staff with which to conduct research protocols*	• *conservative medical community ; PET's clinical utility not demonstrated* • *high costs; low reimbursement*
	Incentives to Use	• *centers of excellence in epilepsy treatment and cardiothoracic surgery* • *growing interest in oncology*	• *scanner located at VAMC favorable for VA patients* • *VAMC supportive of PET center*	• *PET director's main focus is clinical PET applications*	• *VAMC respected in medical community* • *PET director highly regarded* • *VAMC supportive of PET center*
K	Barriers to Use	• *lack of reimbursement for more recently developed clinical oncology applications*	• *location of scanner not favorable to non-VA patients because of the following perceptions:* - *PET unavailable to non-VA patients* - *PET viewed as only a research tool* - *quality of care at VA is poor*		• *potential competition from second PET center in area* • *conservative medical community; clinical utility not demonstrated* • *centralized decision making in CO undermines VAMC's ability to adapt to local market changes* • *viability of VA system in question; creates morale problems*
	Incentives to Use	• *rapidly expanding interest in oncology* • *diverse use of PET encouraged by PET director* • *consensus building approach facilitates likelihood of reimbursement*		• *funded research focused in neuroscience* • *approval for reimbursement of clinical studies reached by consensus*	• *VA administration supports PET's role in the tertiary care setting* • *expertise of PET director highly regarded* • *approval for reimbursement of clinical studies reached by consensus*

SP=Sharing Partner

Table 14: Recommendations Volunteered During VHA PET Site Visit Interviews

Note: These recommendations were not part of the formal interview questionnaire, but were offered by some interview subjects during the interview process. Direct quotes are noted; other recommendations are paraphrased based on information obtained from interview summaries.

Recommendations (frequency)
I. Recommendations for/Comments on Improving VA Systemwide
Examples: • *VA must compete more aggressively in the managed care environment if it is to survive.(8)* • *VA should sponsor more MR and/or CT technology centers and conduct advanced studies to become state-of-the-art in these technologies, rather than invest in PET. (4)* • *High technology should be located at regional facilities.(3)* • *VA needs to focus on support facilities (eg. parking), rather than on PET. (1)* • *“Someone in a responsible position needs to review and better prioritize the allocation of funds for and within the VA system.” (1)*
II. Recommendations for Improving VHA PET Activity
Examples: • *Each PET program should be reviewed critically in terms of its viability, capacity, and available expertise.(3)* • *“VA should stop supporting university PET centers who don’t reciprocate.”(2)* • *“There should be a working cooperative group to decide what needs to be done, and the equipment also must be standardized.”(1)* • *PET should be judiciously placed throughout the system.(1)* • *A strong multi-disciplinary team is needed to run a PET Center. (2)* • *VA should build PET teams similar to a GRECC (Geriatric Research Education & Clinical Center) at a few locations of high expertise with a steady flow of funds to support these teams to carry out a wide range of studies. (1)* • *“VA should have mandated and funded the PET centers. They (VA) weren’t organized as a group to do anything. You’ve got to set it up to make it work.” (1)* • *“You need a paid staff, not just graduate students, making the (cyclotron) materials...” (1)* • *The VA system should create a PET “roving maintenance” team to service all of the PET centers. (3)* • *VA should create a central warehouse for scanner parts and consider a group purchase for upgrade capabilities, using its economy of scale advantage. (1)* • *PET directors need to promote PET’s capabilities more. (2)* • *“VA should invest in PET, especially in neurology and psychiatry.” (1)* • *When purchasing new equipment (i.e. PET camera), you should obtain the largest field of view possible.” (1)*

Recommendations (frequency)
III. Recommendations for Improving PET Sharing Agreements
Examples: • *VA should pursue shared procurement of equipment, start-up funding, and the building as part of the overall initial plan. (1)* • *Strategic planning, which includes “marrying” capital procurement and plant construction, is essential. (3)* • *New sharing agreements should make provisions for equipment upgrades. (2)* • *Sharing agreement negotiations should include, but go no higher than, the office of the regional director to insure that regional needs are considered and duplication within the region is avoided. (1)* • *PET directors with dual appointments should be allowed to participate in sharing agreement negotiations. (1)*
IV. Recommendations for PET Research
Examples: • *VA multi-center studies of the efficacy and cost effectiveness of PET should be done. (15)*   • *The following areas of interest were cited: comparing Thallium reinjection to PET in myocardial viability determination, neurotransmitters, and oncology applications such as ENT, breast, gliomas, solitary pulmonary nodules, and colorectal cancers.* • *VA should support existing centers, but not expand, and use its resources to evaluate the clinical utility of PET and improve its technical capabilities. (7)* • *“The government and third parties must get together to do the research needed to establish the effectiveness of various technologies and fund ways of paying for them.” (1)* • *“If VA is going to invest in high technology, they should be more oriented to academic research.” (1)* • *VA needs one research consultation center. (1)* • *“VA shouldn’t buy more PET scanners for research, but maybe for clinical studies.” (1)* • *A comparison between magnetoencephalography (MEG) and PET is needed, because currently MEG has better temporal resolution. (1)* • *“My tax dollars should not pay for clinical PET neurology/psychiatry applications, only research.” (2)* • *A PET center should have available hardware for brain research. (1)* • *VA should explore the use of PET in neurodegenerative diseases and evaluating brain tumors. (1)* • *A prospective study comparing SPECT, Thallium and PET with technetium in brain tumors should be done. (1)* • *PET should be compared with Thallium in the detection of vasomotor ischemia. (1)*

Recommendations (frequency)
IV. Recommendations for PET Research (continued)
Examples: • *The future of psychiatric research should involve regional development of ligands so to avoid duplication. (1)* • *VA should invest in PET cardiovascular clinical applications and research. (1)* • *Rapid sequence MRI might be compared with PET in a randomized controlled trial to assess effectiveness in the determination of myocardial viability. (1)* • *Future research should look at using PET to detect the site of the unknown primary tumor. (1)* • *A comparison of PET to Gallium scanning in the detection of residual disease in the treatment of Hodgkin's or non-Hodgkin's disease is needed. (1)* • *VA should invest in cancer studies, in one or two well-funded areas. (1)* • *A clinical trial comparing PET with surgical staging is needed. (1)*
V. Other Recommendations/Comments
• *People need to be educated on the limitations of the (PET) technology. (1)* • *"PET is a technology looking for an application." (3)*

Table 15: Best Practices Identified at VHA PET Sites

Issues Addressed	Approach/Process Description	Site/ Contact Person/ Phone Number
Facilitation of the sharing agreement process	*This site has successfully negotiated several sharing agreements, because the details of the sharing agreements are negotiated primarily by the Director's Office, not by Supply Service. However, approval of these sharing agreements does require concurrence by Supply Service. Underlying the success of these negotiations is the trust and respect which have been developed between the partners over time. Negotiating through the Director's Office offers the following advantages:* *1) the Director's Office carries more weight than Supply Service in the negotiation process;* *2) the Director's Office emphasizes involvement in the whole process, including costs, services exchanged, and personnel involved;* *3) the Director's Office can negotiate details with greater flexibility than Supply Service.*	*San Antonio, TX/* *Louise Parker/* *(210) 617-5220*
Facilitation of the protocol approval process	*The Research Imaging Center (RIC) developed a process, based on an NIH model, whereby PET protocols are typically reviewed and approved within 3-4 weeks from the time of submission. To facilitate this process, the following preliminary steps were developed:* *1) The RIC works closely with other committees such as the Radiation Drug Research Committee, Radiation Safety Committee, and the Investigational Review Board to develop mutually acceptable terminology and the forms necessary to facilitate the approval process;* *2) Potential investigators are encouraged to attend weekly PET "lab" meetings to discuss informally their ideas and obtain feedback from PET experts;* *3) Potential investigators are advised to identify a <u>sponsor</u>, who is a member of the core RIC staff, to act as a mentor. As a mentor, this staff member must have the appropriate expertise in the chosen area of study and be familiar with the protocol approval process to assist with protocol development, which may take several months. Additionally, the mentor agrees to vouch for the integrity of the protocol.* *The investigator proceeds to the <u>Protocol Review Committee</u>, comprised of experts in all related RIC disciplines, for review and approval of his or her PET protocol. All sharing partners are represented, although the focus of representation is interdisciplinary, not institutional. A <u>Scientific Advisory Board</u> has been created whose roles are to monitor the Protocol Review Committee for fairness and to advise them in the direction of studies of particular interest to their respective institutions. Each institution is represented equally by members who report to their facility's director.*	*San Antonio, TX/* *Tuhin Chaudhuri, MD/* *(210) 617-5117* *or* *Peter Fox, MD/* *(210) 567-8150*
Standardization of a method used to measure resources directly utilized in a given PET protocol	*Each approved PET protocol, whether for clinical or investigational purposes, is assigned a <u>relative value unit (rvu)</u> which is used to measure directly the resources utilized in that protocol. A rvu consists of the ratio of personnel services utilized (expressed in hours) to commodities and supplies (at actual cost) utilized and is then compared to that of a definitional unit.* *The specific cost for a given protocol is determined by an assessment of overall programmatic costs divided by the expected program value (in rvu's) and then multiplied by the specific protocol rvu. The assignment of a rvu to a given protocol is subject to the review and concurrence of representatives of both sharing partners and is part of the negotiated sharing agreement process.*	*Ann Arbor, MI/* *Milton Gross, MD/* *(700) 374-7886*
Reimbursement by local payers	*Using group consensus, this site created the <u>Review Council for Clinical PET</u> comprised of the clinical PET community from participating medical facilities, the University, local third party payers, and the Health System Agency from NY State (local regulatory bodies) whose goals are:* *1) to develop consensus on acceptable clinical PET protocols to be reimbursed, and* *2) to authorize physicians who have been trained in PET to read clinical PET scans.*	*Buffalo, NY/* *Jayakumari Gona, MD* *or Alan Lockwood, MD/* *(716) 862-3635*
Reimbursement by local payers Expansion of referral base Determination of the clinical usefulness of PET	*To permit eligibility for third party reimbursement, the state required the PET center to develop <u>demonstration protocols</u> to accomplish the following:* *1) to collect data to determine both the clinical usefulness of PET imaging and the cost impact by comparing the costs of treatment plans of referring physicians before and after PET scanning, and* *2) to allow eligible citizens of the state of Connecticut access to PET imaging.*	*West Haven, CT/* *Robert Soufer, MD/* *(203) 937-3427*

Table 16: Research Activity at VHA PET Sites as of October 1994

Note: Information from Ann Arbor and Pittsburgh was obtained through site visit interviews. All other site information was obtained from the pre-site visit surveys.

Abbreviations are listed at the end of the table.

Description	Site*	Funded (X=yes)	Activity			Study Subject		Comments
			Past	Current	Future	Human	Animal	
NEUROLOGY								
Cognitive disorders								
Psychiatric symptoms on cortical metabolism in Alzheimer's disease	WLA	X		X		X		
PET in people at risk for familial Alzheimer's disease	WLA	X		X		X		
Attention deficit and central executive discontrol in Alzheimer's disease	WLA				X	X		
Prodromal Alzheimer's disease	IND	?		X		X		
Familial Alzheimer's Disease	IND	?		X		X		
Prediction of Alzheimer's disease in a 2 year follow up study	AA	?		X		X		
Exploring the diagnosis and treatment of Alzheimer's disease	PITT	X		X		X		
Measurement of regional cerebral blood flow in patients with known or suspected AIDS dementia complex	MINN	X		X		X		
Measurement of rCMRglc in subjects with known or suspected AIDS dementia complex using F-18-FDG and PET	MINN	X		X		X`		
Differential diagnosis of early dementia	AA	?		X		X		
Neural correlates of visuospatial working memory	MINN							info not provided by investigator
PET and reaction time studies of language processing using O-15 water	MINN							info not provided by investigator
Functional neuroanatomy of human cognition using O-15 water	MINN							info not provided by investigator
PET studies of language function	BUF	X		X		X		
PET studies of hearing loss and tinnitus	BUF	X		X		X		
Neurophysiology of pain	AA	?		X		X		
Motor disorders								
The transplantation of fetal substantia nigra into the caudate nucleus and putamenal nucleus of patients with Parkinsons disease	WH			X		X		

Description	Site*	Funded (X=yes)	Activity			Study Subject		Comments
			Past	Current	Future	Human	Animal	
PET studies of Parkinson's disease	BUF				X	X		approved, not funded
Dopaminergic PET and motor dysfunction in Parkinsonism	MAD	X		X		X		
Measurement of rCMRglc in subjects with known or suspected hereditary or sporadic/acquired ataxia using F-18-FDG and PET	MINN			X		X		
Measurement of rCMRglc in subjects with extrapyramidal movement disorders using F-18-FDG and PET	MINN			X		X		
Measurement of regional cerebral blood flow in patients with ataxia	MINN			X		X		
Functional brain mapping in adults with infantile hemiplegia: A PET study of cerebral plasticity	SA	X		X		X		
Investigation of the neural bases of chronic stuttering	SA			X		X		
NIH program project: stuttering, a movement disorder	BUF	X			X	X		
Studies of non-catecholic l-Dopa analogs	MAD	X		X		X		
PET probes of dopamine neurons in young and aged macaques	MAD	X		X		X		
Epilepsy								
Studies of brain blood flow and metabolic function in epilepsy	WH			X		X		
Collaborative interictal PET imaging of epileptic patients	WH			X		X		
Regional cerebral blood flow and glucose metabolism in patients with complex partial seizures	SA			X		X		
The role of PET in predicting outcome following anterior temporal lobectomy for medically refractory partial complex seizures	PA			X		X		
Pre-frontal dysfunction in frontal lobe epilepsy	WLA	X		X		X		
Functional mapping of the brain to monitor blood flow in epilepsy patients who follow a research paradigm involving naming objects	AA	?		X		X		
Other								
Use of PET imaging for the early detection of malignant degeneration of low grade gliomas	WH			X		X		
Measurement of rCMRglc in subjects with chronic cocainism using F-18 fluorodeoxyglucose and PET	MINN			X		X		
FDG PET imaging of cocaine infusion	WLA	X		X		X		
Measurement of rCMRglc in normal volunteer subjects using F-18-FDG and PET	MINN			X		X		

Description	Site*	Funded (X=yes)	Activity			Study Subject		Comments
			Past	Current	Future	Human	Animal	
Measurement of regional cerebral blood flow in normal volunteer subjects	MINN			X		X		
Activation studies of the normal human frontal lobe	WLA	X		X		X		
Auditory activation	IND	?		X		X		
rCBF activation	IND	?		X		X		
rCBF activation	IND	?		X		X		
$rCMR_{glu}$ control studies	IND	?		X		X		
Human Functional Brain Mapping with PET: Inter-subject variability	SA			X		X		
Human Functional Brain Mapping: Brain representation of body schema	SA			X		X		
Use of high brain/blood partition coefficient inert diffusible blood flow tracers in the detection of local blood flow changes	SA			X		X		
Functional and structural imaging in closed-head trauma	SA	X		X		X		
PET studies of minimal traumatic brain injury	BUF	X		X		X		
O-15 peripheral vascular studies in patients with spinal cord injury	WLA				X	X		
PET studies of hepatic encephalopathy	BUF	X		X		X		
PET and neuropsychological studies of cerebral function in patients with chronic severe ischemic coronary artery disease	BUF				X	X		approved, not funded
Mental function in aging	WLA				X	X		
O-15 cerebral activation studies in patients with Persian Gulf Syndrome	WLA				X	X		
Discordant twins	IND	?		X		X		
GSS Indiana kindred	IND	?		X		X		?? description
Action of morphine in the brain	AA	?		X		X		
PSYCHIATRY								
Mood disorders								
Fluoxetine effects on mood, cognition and metabolism	SA	X		X		X		
Functional neuroanatomy of mood, brain glucose metabolism in idiopathic depression and depression associated with basal ganglia disorders	SA	X		X		X		
Effect of prozac treatment on mood, cognition and brain glucose metabolism in patients with primary unipolar depression	SA	X		X		X		
Functional neuroanatomy of emotion: A PET Brain Mapping study	SA			X		X		

Description	Site*	Funded (X=yes)	Activity			Study Subject		Comments
			Past	Current	Future	Human	Animal	
Affect, depression, and brain asymmetry	MAD	X		X		X		
Affective style: Social and psychobiological substrates	MAD	X		X		X		
Exploring the diagnosis and treatment of depression	PITT	X		X		X		
Evaluating the role of the serotonin system in antidepressant therapy	PITT	X		X		X		
Anxiety/Stress disorders								
PET measurement of benzodiazepine receptors in stress	WH	X			X	X		
PET and SPECT measurement of the benzodiazepine receptor in anxiety	WH	X		X		?		
PET measurement of the benzodiadepine receptorwith C-11 iomazenil in patients with anxiety disorders and healthy subjects	WH	X		X		X		
Neurobehavioral correlates of PTSD symptoms in combat veterans	MINN	?				X		info not provided by investigator
PET measurement of cerebral metabolic correlates of yohimbine administration in PTSD and healthy controls	WH			X		X		
CNS activation during episodes of mental stress induced myocardial ischemia	WH	X		X		X		
PET evaluation of treatment for simple phobia	PA		X			X		
Other								
Exploring the diagnosis and treatment of schizophrenia	PITT	X		X		X		
Chemical exposure	IND	?		X		X		? description
Human biological clock	IND	?		X		X		?description
CARDIOLOGY								
Metabolic effects of chronic myocardial hibernation and reperfusion using FDG and O-15 water	MINN	X	X				X	
Myocardial glucose utilization following cardiac surgery	MINN		X				X	
Imaging myocardial perfusion	SA			X		X		
Imaging myocardial viability	PA,SA, IND	X,X,?		X		X		
Imaging myocardial ischemia	PA,SA,IND	X,X,?		X		X		
Screening of healthy volunteers for possible inclusion in a myocardial PET imaging study	SA			X		X		
Role of PET with FDG in conjunction with maximal exercise stress in the assessment of chronic stable coronary artery disease	SA	X		X		X		

Description	Site*	Funded (X=yes)	Activity			Study Subject		Comments
			Past	Current	Future	Human	Animal	
Studies of cost effectiveness of cardiac diagnostic studies	BUF	X			X	X		
Comparative accuracy of rest-redistribution Thallium SPECT vs. FDG PET in predicting reversibility of left ventricular dysfunction following coronary artery bypass surgery	PA		X			X		
A comparison of Rb-82 PET and TI-201 SPECT in the evaluation of CAD	PA		X			X		
The acute effects of cigarette smoking on myocardial perfusion as evaluated by PET	PA		X			X		
Noninvasive PET imaging of cardiac transplant patients	PA		?			X		
Heart dosimetry of 18-F-FDG	PA				X	X		
Evaluation of ischemic heart disease in women: clinical center	PA	X			X	X		under review
Pathogenesis of symptomatic vs. silent myocardial eschemia	WLA			X		X		
Myocardial perfusion by Cu-60 copper PTSM with PET	MAD	X		X		X		
Indicators of metabolism within a perfusion-viability gradient	MAD	X		X		X		
Measuring women's response to cardiac stress with circulating estrogen to explain false positive thallium stress results and replace cardiac caths with PET	PITT	X		X		X		
Exploring severe heart failure and the use and mechanism of beta-blockers	PITT	X		X		X		
ONCOLOGY								
The role of PET-FDG in detection of occult cervical lymph node metastases	MINN			?		X		
Applications of PET in colorectal carcinoma patients	BUF			X		X		
Use of FDG PET scanning to stage esophageal cancer	BUF			X		X		
Lymph node metastases	IND	?		X		X		
Staging patients with lymphoma using whole body imaging	AA	X		X		X		
Lymph node involvement in patients with malignant melanoma	AA	X		X		X		
Staging of the mediastinum for non small cell lung cancer	AA	X		X		X		
Evaluating solitary pulmonary nodules	AA	?		X		X		
Monitoring chemotherapy in the treatment of breast cancer	AA	X		X		X		
Defining the variable needed to monitor physiologic changes in tumors prior to tumor shrinkage after chemotherapy	PITT	X		X		X		

Description	***Site****	***Funded (X=yes)***	***Activity***			***Study Subject***		***Comments***
			Past	***Current***	***Future***	***Human***	***Animal***	
OTHER								
Studies of dialysis disequilibrium	BUF	X		X		X		
PET studies of inhaled 11-C-triamcinolone	BUF	X		X		X		
Clinical evaluation of the Argus PET system	MAD	X	X			X		
Pancreatic blood flow	IND	?		X		X		
Skeletal muscle	IND	?		X		X		

Abbreviations: AA=Ann Arbor
BUF=Buffalo
IND=Indianapolis
MAD=Madison
MINN=Minneapolis
PA=Palo Alto
PITT=Pittsburgh
SA=San Antonio
WH=West Haven
WLA=West LA

IV. SUMMARY

It has been only until recently that all VHA PET centers have become fully operational, thus allowing for an assessment of their activity. The information from the pre-site visit surveys indicated that there are significant variations in the characteristics of the PET centers and in the types, volumes and purposes of the PET studies across all sites. Information from site visit interviews indicated that there are important organizational, professional, scientific, and reimbursement issues yet to be overcome before PET becomes more widely diffused. Recommendations volunteered by some of the interview subjects including processes by which to overcome some of these barriers (See Table 14) were presented, and may be helpful to administrators, researchers and clinicians.

Although there is a growing interest in clinical PET studies, PET is still viewed by regulators and the general medical community as a research tool. PET has made a significant contribution to overall research activity within VHA and continues to be the primary research tool for certain areas of research, particularly in the neurosciences.

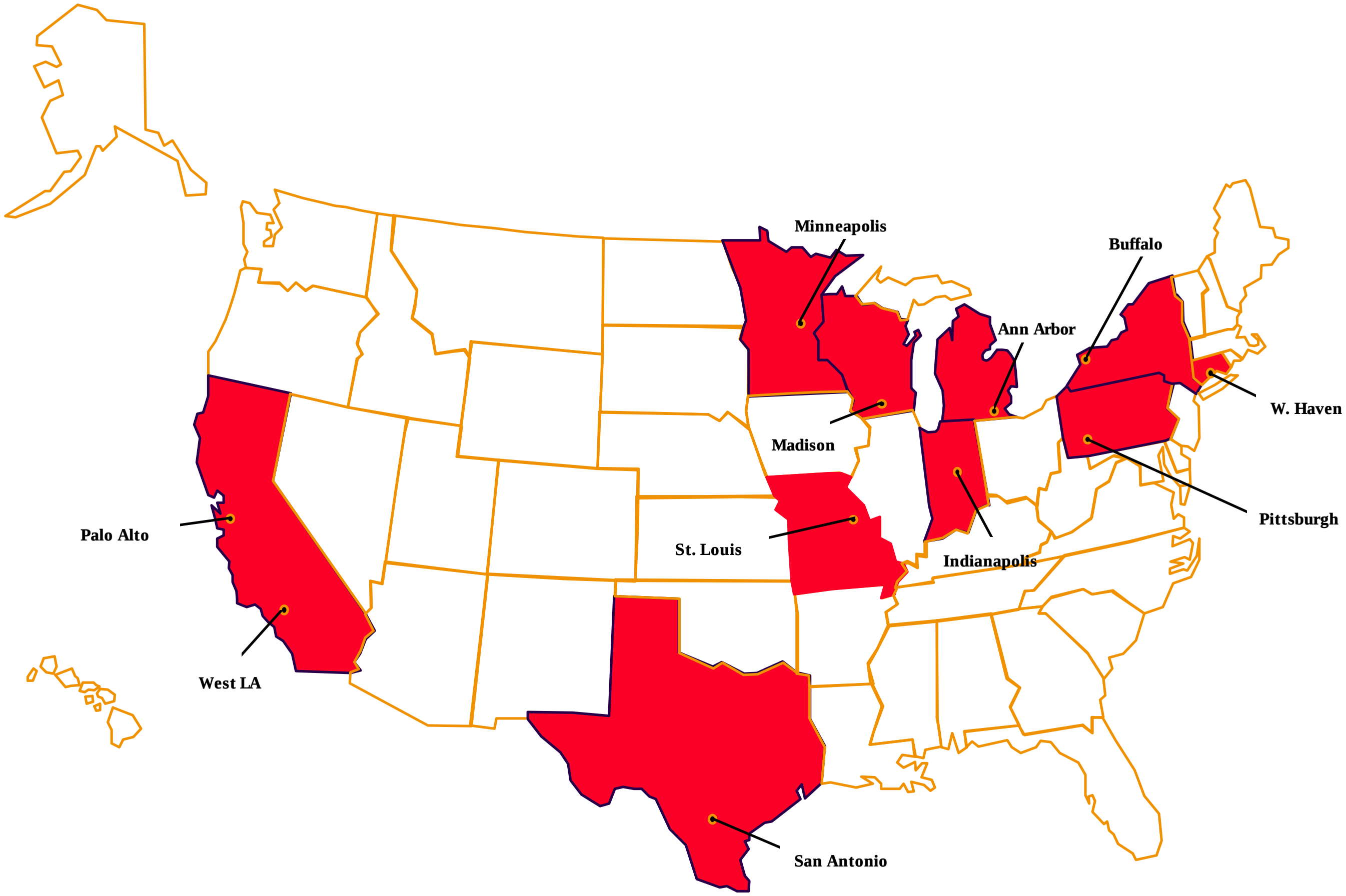


Fig. 1 Locations of VHA PET Centers

Name of PET Center: ____________________________

PRE-SITE VISIT SURVEY

I. GENERAL INFORMATION

1) On average, how many hours/day is the PET Center in operation? _______hours/day
2) On average, how many days/week is the PET Center in operation? _______days/week
3) On average, how many days/week is the PET Center available to VA patients? ____days/week
4) On average, how many weeks/year is the PET Center open? _____weeks/year
5) Is the PET Center operational on federal holidays? ____yes ____ no
6) On average, how many days/year is downtime experienced for:
   _________ scheduled maintenance _________ emergency maintenance
7) Is the PET Center affiliated with an academic institution? ___ yes ___ no If yes, what is the affiliate's name?______________________________________
8) Who is responsible for scheduling patients?___VA ___Affiliate ___Center (freestanding)

II. PET SYSTEMS

1) What year did PET become operational at your site?_______
2) What is the current name and model of your PET camera? ________
3) Has there been an upgrade in equipment since you first became operational? ___ yes ___ no
4) Is the scanner located at VA? ___ yes ___ no If located off-site, how many miles away from your facility is the scanner? ________ miles
5) Does your facility own a cyclotron? ___ yes ___ no (If no, go to #6) If yes:
   a) Where is it located?____
   b) What is the current name, model, and age of the cyclotron? _____________
   c) Which products are generated from this cyclotron?
   d) Do you supply cyclotron-generated radiopharmaceuticals to other PET facilities? __yes __ no If yes, how much revenue was generated from your cyclotron products in FY "92? $_____________ FY '93? $____________ FY '94? $_____________
   e) What mode (s) of transportation (i.e. plane, truck, pneumatic tube, etc.) is/are used to deliver radiopharmaceuticals(s) to your PET facility? _______________ to other PET facilities? _______________

GO TO #7

6) Does your PET Center use cyclotron products? ___ yes ___ no (Go to #7) If yes:
   a) Where are these cyclotron products manufactured? _________
   b) What mode(s) of transportation (i.e. plane, truck, pneumatic tube, etc.) is/are used to deliver radiopharmaceuticals(s) to your PET facility? _____________
7) Does your PET center use generator-produced radiopharmaceuticals for PET? ___ yes ___ no (If no, go to #8) If yes:
   a) Where are these generator products manufactured? _________
   b) What mode(s) of transportation (i.e. plane, truck, pneumatic tube, etc.) is/are used to deliver radiopharmaceuticals(s) to your PET facility? ___________
8) Do you supply generator-produced radiopharmaceuticals to other PET facilities? ___ yes ___ no (If no, go to #9) If yes:
   a) Do you supply cyclotron-generated radiopharmaceuticals to other PET facilities? __yes __no
      If yes, how much revenue was generated from your cyclotron products in
      FY “92? $_____________ FY ‘93? $_____________ FY ‘94? $_____________
   b) What model(s) of transportation (i.e. plane, truck, pneumatic tube, etc.) is/are used to deliver radiopharmaceuticals(s) to your PET facility?________
      to other PET facilities? __________
9) Does your facility have a Rubidium generator? ___ yes ___ no If yes:
   a) What is the name, model and age of the generator? _____________

III. PERSONNEL

1) To help us determine the salary costs attributed directly to PET, please complete the following tables for information on current staff:

Existing Staff Expertise	***#VA FTEEs***	***VA Salary & Benefits***	***% Grant Funded***	***#Non-VA FTEEs***	***Non-VA Salary & Benefits***	***% Grant Funded***
Physician **See chart below	XXX	XXX	XXX	XXX	XXX	XXX
Radiochemist						
Medical Physicist						
Radiopharmacist						
Nuclear Med Tech						
Chem Tech						
RN						
Administrator						
Director(if non-physician)						
Secretary/ Receptionist						
Other:						

Physician Expertise (Give Specialty)	***VA-based Salary & Fringe ($)***	***% Grant Funded***	***% Total VA Time Devoted to PET Center Clinical or Research Applications***	***% Total VA Time Devoted to PET Center Administration***
1.				
2.				
3.				
4.				

2) Are there other positions that need to be filled for successful operation of your PET facility? ___

If yes, please indicate:

opened positions ___________________________

positions to be created __________________

3) Do you provide on-the-job training for your PET Center staff? ___ yes ___ no

4) Do you participate in a formal instructional program designed to train PET personnel? __yes __ no

5) To determine each staff member's function(s) at your PET facility, please check (√) the box(es) that correspond(s) to the appropriate function(s) of each staff member:

Staff Expertise	***Running Cyclotron***	***Eluting Generator***	***Quality Control***	***Radiolabelling Synthesis***	***Administering Dose to Patient***
Physician					
Radiochemist PhD					
Radiochemist Tech					
Medical Physicist					
Radiopharmacist					
Nuclear Med Tech					
Chem Tech					
RN					

IV. RESEARCH

1) Attach a list of past and current research projects performed by your facility involving PET since FY 92. Indicate whether or not they are funded, source of funding, and any resulting publications.

2) Attach a list of any proposed projects in PET application areas that are planned for your facility and/or with your university affiliates.

V. SPACE REQUIREMENTS

1) To determine issues that may affect the use of PET, please indicate how much space is currently allocated for:

Cyclotron ____________ sq. ft. Is this adequate? _____ yes _____ no

PET camera ____________ sq. ft. Is this adequate? _____ yes _____ no

Electronics/control room ____________ sq. ft. Is this adequate? _____ yes _____ no

Radiochemistry Lab ____________ sq. ft. Is this adequate? _____ yes _____ no

Shop facilities ____________ sq. ft. Is this adequate? _____ yes _____ no

Administration ____________ sq. ft. Is this adequate? _____ yes _____ no

Waiting area ____________ sq. ft. Is this adequate? _____ yes _____ no

parking spaces for PET Center ____________ sq. ft. Is this adequate? ____yes ____ no

VI. ANCILLARY SERVICES

1) To identify the difference(s) between PET sites with respect to potential referral sources, please indicate whether or not these services exist at your VA or University affiliate:

Service	***Yes***	***No***	***Service***	***Yes***	***No***
Alcohol Dependency Treatment Unit			Neuropsychological Testing		
Cardiac Cath Lab			Nursing Home Care Unit		
Cardiac ICU			Patient Health Ed. Program		
Cardiac Surgery Program			PFT Lab		
Electron Microscopy			Sickle Cel l Screening Program		
Hemodialysis and CAPD Trainig			Speech Pathology Lab		
Hypertension Screening and Treatment Program			Surgical ICU		
Medical ICU			Geriatric Research Education & clinical Center (GRECC)		
Mental Hygiene Clinic			Women's Health Center		
PTSD Program			Health Psychology Program		
Epilepsy Program			Cancer Center		
Other:			Other:		

2) To identify the potential referral base of each PET facility, please fill in the following table and be as complete as possible:

Specialty	***Affiliated Residency or Fellowship Program at VA?***		****Potential Referring **VA Physicians in Each Specialty***	***Number of **VA Physicians Who Have Referred***	****Potential Referring Non-VA Physicians in Each Specialty***	***Number of Non-VA Physicians Who Have Referred***
	Yes	***No***				
Cardiology						
Oncology						
Neurology						
Psychiatry						
Pulmonary						
Internal Medicine						
ENT						
Oncology						
Gynecology						
GI						
Other:						
Surgical:						
Cardiac						
Neuro						
Other:						

*Note: "Potential Referring" physician is defined as a Physician who would refer patients from his/her clinical practice, not onw who is strictly a researcher.

**Note: "VA physician" is defined as a physician who is employed by VA for 5/8 time or greater. This information may be available through your Chief of Staff.

VII. BUDGET

1) To identify costs attributed directly to PET, please fill in the following table:

Item	*Total FY '92 Costs*	*Total FY '93 Costs*	*Total FY '94 Costs*
FIXED SUPPLIES:			
Cyclotron			
Generator-Related			
Maintenance Contract for Cyclotron			
Maintenance Contract for Camera			
Insurance			
Other:			
VARIABLE SUPPLIES:			
Film			
Purchased Radiopharmaceuticals**			
Other Pharmaceuticals i.e. Persantine, Adenosine			
Cyclotron Supplies Including Target Materials			
Patient Supplies			
Camera-Related Supplies Including Rod Source			
Other:			

**If a non-cyclotron or non-generator site

VIII. OTHER

1) What is your definition of <u>clinical</u> PET?

Name of PET Center: ___________________________

Interview Questionnaire- PET Chiefs

1. How long have you been at this VA facility?
2. What is your current title? previous title?
3. Were you involved in the planning of this PET facility?
    a. Who else was involved?
    b. Who made the decision as to whether or not PET would be available at this facility?
4. Could you explain your facility's PET Sharing Agreement?
5. Does the availability today of PET for VA patients differ from expectations specified in the Sharing Agreement?____ yes _____no
    a. If yes, in what ways?
    b. Why do you think this is the case?
6. With respect to planning for PET, if a sharing agreement were renegotiated, what would you do differently?
7. What does having access to PET technology mean to this facility? (financial implications, status, etc.)
8. What kinds of financial and administrative support have been provided for this PET facility, i.e.:
    a. Was space provided?
    b. Was a building provided?
    c. Was start-up funding provided?
    d. Who gets third party revenue?
    e. Provisions for marketing?
    f. Others?
9. Have you had difficulties obtaining reimbursement for PET scans? Explain.
10. What barriers can you think of that affect the use of PET?
    a. What has the VAMC done to contribute to, eliminate or reduce these barriers?
11. Does this facility have MR capabilities? CT? SPECT?
    a. What generation is the equipment?
    b. What are its capabilities?
    c. Has it impacted the use of PET? If so, how?
12. Hypothetical: If you were starting from scratch a could afford to buy only one state-of-the-art imaging technology, which one would you buy and why?

13. Where do you see this VAMC going 3-5 years down the road with respect to the managed care environment?
    a. What do you see as PET's role in this?

THE FOLLOWING QUESTIONS ARE BASED ON THE PRE-SITE VISIT QUESTIONNAIRE:

14. Are there problems scheduling VA patients for PET scanning?___yes ___no
    a. What are those problems?
15. What percentage of scans are inpatients? outpatients?
16. Where is the closest PET facility?
    a. How long does it take to get there?______minutes
17. Are there any other geographic factors that affect access?
    a. If yes, what are they?
18. How many personnel do you have?
19. Did you experience difficulties in recruiting personnel for PET?____yes ___no
    a. If yes, please explain:
20. Do you currently have any vacancies? ____yes ____no
    a. If yes, what positions are vacant?
21. How will staff expertise be recruited to these new positions?
22. (See Section III #3) If on-the-job training is provided, describe what kinds of training is provided**:**
23. (See Section III #3) If a formal instructional program is provided, describe what kinds of training is provided**:**
24. Please list any workshops, presentations, grand rounds, etc. given as an effort to educate and inform the medical staff at your facility of PET:
25. Have the efforts listed above resulted in a change in the number of referrals to your PET facility?____yes ____ no.
    a. If yes, please describe in terms of volume and types of scans requested:
26. How active is the affiliated university medical center in the PET Center in terms of the proportion of time and equipment used?
27. Do you have other collaborative efforts with other institutions, facilities or providers? _____ yes _____ no.
    a. If yes, please describe them:
28. Are there opportunities for sharing resources beyond what your program is doing?
    a. Please describe what they might be:

Name of PET Center: __________________

Specialty: _____________

Interview Questionnaire- Physicians

1 Are you employed by VA? _____yes _____no
   a. If yes, what percentage of time is devoted to VA?
   b. Do you attend at the University Hospital? _____yes _____no
2. Are you employed by the University? _____yes _____no
   a. If yes, what percentage of time is devoted to the University?
   b. Do you attend at the VAMC? _____yes _____no
3. How long have you worked for: VA?_____yrs University?_____yrs
4. Is your interest primarily clinical, research or both?
   a. How is your time divided between clinical, research, administration and other duties?
5. Do you have a lab? _____yes _____no
6. Do you see patients? _____yes _____no
7. How did you first learn about PET technology?
8. How did you first learn about PET at your facility?
9. Do you refer patients for PET scans? _____yes _____no
10. If you do not refer, why not?
11. If you do refer, for what conditions? For each condition give the following information:
    a. Is this a research protocol and/or is this used for clinical purposes?
    b. What tests are ordered prior to your ordering a PET scan?
    c. In this situation, what does PET give you that the other tests or technologies do not?
    d. If PET is ordered for clinical purposes, for what percentage of patients with this condition is PET ordered?
    e. Can you also give absolute numbers?
12. Have you experienced problems scheduling patients for PET?
    a. If so, what are these problems?
13. Do your colleagues believe in PET?
    a. How many refer vs. how many do not refer?
14. Are there any other issues to discuss?

Appendix 10

Assessments, Guidelines, and Policy Statements Produced by Other Agencies

Author: Karen Flynn, D.D.S., M.S., *Manager, MDRC Technology Assessment Program*

Appendix 10:
Assessments, guidelines, and policy statements produced by other agencies

As it collected information for this assessment, the MDRC Technology Assessment Program identified the following documents produced by other technology assessment agencies. The findings of assessments in the public domain, or otherwise available to the MDRC, are tabulated.

Date (Reference)	**Issuing Agency**	**Topic/Title**	**Methods**	**Findings/Comments**
1984 (McKhann, et al.)	Work Group on the Diagnosis of Alzheimer's Disease *National Institute of Neurological and Communicative Disorders and Stroke (NINCDS) and Alzheimer's Disease and Related Disorders Association (ADRDA)*	Clinical criteria for the diagnosis of Alzheimer's disease, with statements on the utility of laboratory and imaging assessments (including PET)	Expert panel consensus	• Most patients with Alzheimer's disease show cerebral hypometabolism (on PET imaging) when compared with normal age-matched controls. These changes correlate with disease severity and may be correlated with neuro-psychological test performance. • Since PET reveals a significant variation even among normal subjects, any changes may have to be severe to be detected. • The value of PET studies in determining the stage of disease, in documenting progression, and in assessing the effects of treatment is unknown.
1986	American College of Cardiology *Task Force on Assessment of Diagnostic and Therapeutic Cardiovascular Procedures (with American Heart Association)*	Guidelines for clinical use of cardiac radionuclide imaging	Expert panel	Outdated
1987	American Hospital Association	Positron Emission Tomography (PET): Progress and Opportunities	Narrative review, opinion	Opinions on potential uses and role in hospital presented.
1988	American College of Nuclear Physicians/Society of Nuclear Medicine	Positron Emission Tomography: Clinical Status in US, 1987	Qualitative review and expert opinion	• PET provides unique clinical information in several conditions: - identifying patients with coronary artery disease - predicting response to coronary revascularization procedures - localization of focus in partial epilepsy - management of patients with gliomas - distinguishing recurrent glioma from radiation necrosis • additional applications will be documented in future
1988	American Medical Association, Council on Scientific Affairs, PET Panel	Positron Emission Tomography in Oncology	Qualitative review and expert opinion	• informational report to annual meeting • describes current and potential uses of PET in clinical medicine and research related to oncology: - metabolism and physiology of tumors and effects on adjacent tissues - specific probes for target sites on tumors - quantitative measurement of tumor biology and response to treatment • most oncology applications to date had been qualitative

Date (Reference)	Issuing Agency	Topic/Title	Methods	Findings/Comments
1988	American Medical Association, Council on Scientific Affairs, PET Panel	Application of Positron Emission Tomography in the Heart	Qualitative review and expert opinion	• informational report to annual meeting • describes applications of PET to heart: - blood flow - metabolism - experimental injury • potential for: - investigation of cardiomyopathies - study of neural control of heart - evaluation of effects of drugs on cardiac tissues
1988	American Medical Association, Council on Scientific Affairs, PET Panel	Positron Emission Tomography - A New Approach to Brain Chemistry	Qualitative review and expert opinion	• informational report to annual meeting • describes applications of PET in stroke, epilepsy, malignancies, dementias, and schizophrenia and in basic studies of synaptic neurotransmission
1990	American College of Cardiology	Positron emission tomography	Qualitative review and expert opinion	• myocardial viability - PET is an important clinical research tool - appears to provide unique clinical information - imaging blood flow and metabolism appears to be useful as a diagnostic procedure in selected situations • myocardial perfusion - clinical research procedure rather than routine clinical diagnostic procedure
1990	Australian Institute of Health and Welfare	Positron emission tomography.	Narrative review with cost analysis	Sufficient case has not yet been established for routine use of PET as a clinical service in Australia. If proposed PET units are introduced into Australia, they should be subject to a coordinated evaluation of clinical and cost benefits. No further units should be considered until the evaluations are completed.
1991	Blue Cross and Blue Shield Association	Positron emission tomography		Available to subscribers only
1991	American Academy of Neurology	Assessment: positron emission tomography	Synthesis of literature and expert opinion	• PET with FDG or ^{15}O labelled compounds is safe and efficacious diagnostic clinical technique • provides unique and/or complementary information to that provided by anatomic imaging • clinical efficacy in: - localization of seizure foci - differential diagnosis of dementia and movement disorders - grading of primary brain tumors - localization of brain tumor biopsy sites - differentiation of recurrent high grade gliomas from radiation necrosis • assessment should act as guide for Academy members until more rigorous scientific assessment becomes available • *published commentary (Powers, et al., 1991): assessment "fell far short of reasonable standards for adequate assessment of this diagnostic procedure..."*

Date (Reference)	Issuing Agency	Topic/Title	Methods	Findings/Comments
1992	American College of Cardiology *(with American Heart Association and Society of Nuclear Medicine)*	Standardization of cardiac tomographic imaging		Technical report, not technology assessment.
1992	Australian Institute of Health and Welfare	Cardiac imaging technologies	Review of national health system data plus qualitative review of literature as basis for discussion paper	• Significant increase in number of investigations performed on patients with suspected heart disease (118% increase in Australia from 1989090 to 1991-92 in use of coronary angiography, nuclear medicine, and ultrasound testing) • Increasing use of echocardiography (multiple tests on same patient) may be primarily for physician reassurance • Little direct evidence that benefits (effect on patient outcome) are commensurate with increase in use of tests • diagnostic test guidelines for CAD should be developed • new cardiac imaging methods (including PET) should be evaluated in relation to existing tests
1992	Blue Cross and Blue Shield Association	Positron emission tomography of the central nervous system		Available to subscribers only
1992	Prudential Insurance Company of America	Positron emission tomography for the detection of coronary artery disease and myocardial viability		Proprietary information
1992	Prudential Insurance Company of America	Positron emission tomography for the detection of neural abnormalities		Proprietary information
1993	Swiss Institute for Public Health	PET Consensus Conference - Final Statement		English summary not available.
1993	Swiss Institute for Public Health	PET Information Synthesis		English summary not available.
1994	Blue Cross and Blue Shield Association	Positron Emission Tomography for Assessment of Myocardial Viability		Available to subscribers only.
1994	ECRI	Positron emission tomography for evaluation of ischemic heart disease		Available to subscribers only.
1994	University Hospital Consortium	Positron emission tomography	(qualitative) review and consensus	• Comprehensive assessment addressing economics, clinical uses, regulatory issues, reimbursement issues, and recommendations from perspective of academic medical centers • medical literature supports clinical use of PET in: - cardiology (myocardial viability, perfusion studies before revascularization procedures) - neurology (presurgical localization of epileptogenic foci) - oncology (differentiating necrosis from recurrent brain tumor, breast cancer, colorectal cancer, solitary pulmonary nodules, metastases) • full text available from MDRC Technology Assessment Program

Date (Reference)	Issuing Agency	Topic/Title	Methods	Findings/Comments
1995	Office of Health Technology Assessment	Health Technology Review: Myocardial Perfusion Imaging with Rubidium ^{82} Positron Emission Tomography	(systematic) review	• image of cardiac perfusion produced with ^{82}Rb PET is clearer that with ^{201}Tl SPECT, due to higher pixel counts and attenuation correction • currently available data are insufficient to determine whether improved images lead to higher sensitivity and specificity • planar ^{201}Tl scintigraphy, ^{201}Tl SPECT, and ^{82}Rb PET may be used to evaluate noninvasively most patients referred to a cardiac center • patients with negative scans by any of the methods may not be at increased risk of cardiac event and might be conservatively managed according to clinical condition and symptoms
1995	Agency for Evaluation of Health Care Technologies (Spain)	Positron Emission Tomography in Cardiology	Synthesis of other assessments	English text not available.
1996	Blue Cross and Blue Shield Association	PET Myocardial Perfusion Imaging for the Detection of Coronary Artery Disease	(systematic) review	Available to subscribers only.
1996	Blue Cross and Blue Shield Association	PET Myocardial Perfusion Imaging for the Detection of Coronary Artery Disease - Cost-Effectiveness Analysis	Cost-effectiveness analysis	Available to subscribers only.
In preparation	Medical Research Council *National Health Service, UK*	Evaluation of PET scanning (neurology, oncology)		

References

Agencia de Evaluación de Tecnologías Sanitarias. Tomografía por emisión de positrones en cardiología. Madrid, 1995.

American Academy of Neurology, Therapeutics and Technology Assessment Subcommittee: Assessment: Positron Emission Tomography. Neurology 1991; 41:163-7.

American College of Nuclear Physicians/Society of Nuclear Medicine, Task Force on Clinical PET: Positron Emission Tomography: Clinical Status in the United States in 1987. J Nucl Med 1988; 29:1136-43.

American Hospital Association. Positron emission tomography (PET): progress and opportunities. 1987.

American Medical Association, Council on Scientific Affairs, Positron Emission Tomography Panel: Positron Emission Tomography in Oncology. JAMA 1988; 259:2126-31..

American Medical Association, Council on Scientific Affairs, Positron Emission Tomography Panel: Application of Positron Emission Tomography in the Heart. JAMA 1988; 259:2438-45.

American Medical Association, Council on Scientific Affairs, Positron Emission Tomography Panel: Positron Emission Tomography - A New Approach to Brain Chemistry. JAMA 1988; 260:2704-10.

Australian Institute of Health and Welfare, National Health Technology Advisory Panel.. Positron emission tomography. 1990.

Australian Institute of Health and Welfare. Cardiac imaging technologies: a discussion paper. 1992.

ECRI: *Heath Care Standards, 1995*.

Klinik und Poliklinik für Nuklearmedizin, Universitätsspital Zürich und Schweizerisches Institut für das Gesundheitswesen. Poistronen-emissions-tomographie (PET): Konsensus-konferenz schlusserklärung. Zürich/Aarau, 1993.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., and Stadlan, E.M.: Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology 1984; 34:939-44.

Powers, W.J., Berg, L., Perlmutter, J.S., and Raichle, M.E.: Technology assessment revisited: Does positron emission tomography have proven clinical efficacy? Neurology 1991; 41:1339-40.

Report of the NHS Health Technology Assessment Programme 1995.

Schweizerisches Institut für das Gesundheitswesen. Positronen-emissions-tomographie (PET): informationssynthese z.H. des Bundesamtes für Sozialversicherungen zur Unterstützung des Antrages betreffend die Anerkennung von PET als pflichtleistung der sozialen versicherungen. Aarau, 1993.